Rhythm in late-modern Stockholm

Social stratification and stylistic variation in the speech of men

Nathan Joel Young

PhD thesis

Department of Linguistics
Queen Mary, University of London

2019

Declaration

I, Nathan Joel Young, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature: Nathan Joel Young

Date: July 10, 2019

Details of collaboration: The Forced Aligner of Swedish SweFA was built in collaboration with Michael McGarrah at the Georgia Institute of Technology

Details of publications:

Young, N. (2018). Rhythm in Stockholm's two working-class varieties: Separate models predict intervocalic durational contrast. In *Proceedings of 9th International Conference on Speech Prosody, 2018*, 448–452. https://www.isca-speech.org/archive/SpeechProsody_2018/pdfs/103.pdf

Abstract

The main finding of this thesis is that rhythm is a stratified variable in the speech of men in Stockholm. An epicenter for the social forces associated with late modernity, Stockholm is also home to Europe's 'first' multiethnolect (*Rinkeby Swedish*, Kotsinas 1988a). Swedish-language researchers describe the variety as 'staccato', but rhythm has not been thoroughly investigated for *any* variety of Stockholm Swedish in production.

Data come from 36 male Stockholmers, ages 24–45, from a stratified sample of social classes. Seventeen self-identify by the term *svensk* (Swedish) and 19 by the highly racialized term *invandrare* (literal translation: immigrant). All were born in Sweden save for three who arrived before age seven. Three contextual styles were elicited to capture a speech-formality cline: CASUAL, READING, and RADIO (reading like a radio announcer). Rhythm is operationalized with an adaptation of the *nPVIV* algorithm (Low, Grabe, & Nolan 2000).

Not only does rhythm stratify predictably in the direction of *staccato* (low alternation) for the racialized working class, it also is significantly high-alternation/non-staccato in the speech of the white working class. The former is interpreted to be a feature of multiethnolect; the latter a feature of *Södersnack*, Stockholm's industrial-era working-class variety. The higher classes produce an intermediate degree of rhythm in their casual speech.

Rhythm variation among the working classes is also stylistically sensitive. Working-class READING and RADIO appear to target upper-class CASUAL. The racialized working class shows a stylistic sensitivity that is stronger among younger speakers than old, implying a transition from *indicator* to *marker* (Labov 1972a:179) for staccato rhythm. The white working class shows a high degree of stylistic sensitivity regardless of age, implying that high alternation is a *Södersnack* legacy feature.

Generational differences in rhythm production are examined within the racialized working class, and a change point is identified between those born before 1983 and those after. Those born before 1983 mainly achieve 'staccato' with a reduction of accented phonologically-long vowels. Those born after 1983 achieve it with an innovation; they enlarge unstressed vowels, both phonologically short AND long. 'Reduction' and 'enlargement' refer to duration, f_0, and energy. The change point coincides with historical spikes in migration, inequality, and school segregation that would have occurred when the speakers were in adolescence.

In ALL contextual styles, age is a stable predictor of rhythm, independent of social class and racialization. Younger speakers of any background have more staccato speech than older speakers of the same background. It is proposed that this is due to the diffusion of contact prosody, for which multiethnolect is one key conduit.

Sammanfattning

Denna avhandling visar att talrytm är socialt stratifierad i vardagsspråket hos vuxna män i Stockholm. Som ett symboliskt centrum för många senmoderna sociala förändringar i Europa är Stockholm också en hemvist för Europas "första" multietnolekt (*rinkebysvenska*, Kotsinas 1988a). Forskare i svenska har beskrivit varietetens prosodi som 'staccato' och 'stötig' men det saknas fortfarande en grundlig utredning av talrytm för *alla* stockholmska varieteter, inte minst för Stockholms multietnolekt.

Avhandlingens material består av ljudinspelningar med 36 män från Stockholm, i åldern 24–45, från ett stratifierat urval av sociala klasser. Sjutton deltagare identifierar sig med begreppet *svensk* och 19 med det rasifierade begreppet *invandrare* (begreppen problematiseras i avhandlingen). Alla deltagare är födda i Sverige förutom tre som invandrade innan de fyllde sju år. Tre talstilar spelades in för att fånga ett formalitetskontinuum: VARDAGSSPRÅK, HÖGLÄSNING och RADIO (högläsning som härmar en radiopratare). Talrytmen mäts med en anpassad version av *nPVIV*-algoritmen (Low, Grabe, & Nolan 2000).

I enlighet med avhandlingens hypotes stratifierar talrytmen åt staccatohållet (låg vokalkontrast) hos den rasifierade arbetarklassen. Talrytmen stratifierar däremot också åt det motsatta hållet ('ickestaccato', dvs. hög vokalkontrast) hos den vitsvenska arbetarklassen. Jag drar slutsatsen att det förstnämnda resultatet är ett drag av Stockholms multietnolekt och det senare ett drag av så kallat *södersnack*, Stockholms traditionella arbetarmål. Överklassen producerar en prosodisk rytm som ligger mittemellan de två arbetarklassernas rytmer.

Den rytmiska variationen hos arbetarklassen är systematisk och känslig för språkregister. Rytmen i båda arbetarklassernas HÖGLÄSNING och RADIO liknar rytmen i överklassens VARDAGSSPRÅK. Hos den rasifierade arbetarklassen beror rytmens stilmässiga känslighet på ålder. Yngre talare dämpar staccatorytmen i högre grad än äldre talare, vilket tyder på att staccatoeffekten har utvecklats från *indikator* (indicator) till *markör* (marker) (Labov 1972a:179). Den vitsvenska arbetarklassen visar en hög grad av stilmässig känslighet, vilket tyder på att den höga vokalkontrasten är ett etablerat drag som möjligtvis härstammar från *södersnack*.

Det finns en generationsskillnad inom den rasifierade arbetarklassen mellan dem som föddes före och efter år 1983. De som föddes före 1983 utför staccatoeffekten med en reduktion av betonade långa vokaler. De som föddes efter 1983 utför staccatoeffekten med följande novation: en utvidgning av obetonade vokaler, både korta OCH långa. Reduktionen och utvidgningen utförs genom duration, f_0 och intensitet. Generationsskillnaden sammanfaller med betydande socialhistoriska utvecklingar som alla uppträdde när informanterna var i tonårsåldern, t.ex. 90-talets migrationsvåg, en ökad samhällelig ojämlikhet och segregation inom skolväsendet.

För ALLA talstilar är ålder en stabil förklaringsvariabel när nPVIV är responsvariabeln, oberoende av social klass, etnicitet eller rasifiering. Yngre talare, oavsett bakgrund, har lägre vokalkontrast än äldre talare. Jag föreslår att det här kan ses som en följd av en diffusion av prosodiska kontaktdrag in i stockholmska, där multietnolekten är huvudfaktorn.

To my grandmothers Connie and Grace

Memories of their tenacity inspired me to stick it through.

Acknowledgements

Numerous individuals and agencies stepped up to help me with this dissertation. Without their generous resources, time, and grace, the project would have stopped before it started.

The British Economic and Social Research Council generously financed my stipend, training and fieldwork. The Sven och Dagmar Salèns stiftelse generously financed much of the transcription work. Thank you to my excellent transcribers, Adam, Isabelle, and Fanny. I would also like to thank Jan, Ali, and Sodexo Meetings for letting me use their spaces.

It is no small ask, particularly in a place as private as Sweden, to conduct in-depth interviews with people about their social lives and daily habits. The project's 36 participants have made a tremendous amount of space for my inquiries, and for that I am in their debt. Some still remain just a text message away, patiently answering my follow-up questions.

I'd like to thank The Centre for Research on Bilingualism at Stockholm University, its faculty, and its director, Lena Ekberg, for hosting me during the data-collection phase. They provided me with 24-hour access, and facilities in which I could conduct interviews. Carla, thank you for handing over your office – I hope I didn't leave it too messy! To the faculty and staff at the Centre, thank you for all your help and for including me in the daily activities of your fabulous department. These include Caroline, David, Gunnar, Josefina, Kari, Kenneth, Linus, Marta, Natalia, Niclas, Patric, and Tatiana.

Several other individuals in Stockholm helped make the data-collection phase of this project possible. Minna, thank you for being the best housemate I could have hoped for, and thank you for helping correct my Swedish. Thank you also for being there for me emotionally through thick and thin. Keder, thank you for facilitating so many interviews through your vast family network and thank you for all the insider knowledge on Stockholm's multiethnic suburbs. Laura, thank you for linking me to Stockholm's private equity circle – your colleagues provided me with fabulous data. Tom, thank you for letting me record and recruit participants in your barber shop – I know it was a royal pain, so I really am grateful. Ray Ray, thanks for all your help and support in collecting data and being my Stockholm go-to.

During the analysis phase of this project, I spent two months training with Erik Thomas at North Carolina State University. He provided me with invaluable training on speech rhythm and prosody as well as helpful methodological insights on phonetic segmentation. I want to thank him for his unbridled generosity. I also want to thank the other faculty and staff there, including Danica, Jeff, Robin, Stephanie, and Walt.

A pivotal component to the success of this project was the construction of a forced aligner for Swedish (SweFA). This would not have been possible without the help of my 'uncle' Mike McGarrah. He was responsible for building the language-agnostic package for FAVE, which housed my first prototype of SweFA, and he spent many nights on Skype with me troubleshooting my perilous journey into the programming world. Thank you, Lynn, for letting him Skype with me all of those late nights and weekends!

It is hard to condense into words the gratitude and admiration I have for my first supervisor, Devyani Sharma. She has been in my corner on so many fronts. She took the painstaking time to help me apply and secure funding before she even really knew me. Once I was enrolled, she caught me up on the sociolinguistic theory that I did not get in my BA and MA. She dedicated hours and hours to reading and correcting my work. Her feedback molded me into the thinker that I am today. And she also had the guts to tell me whenever I did not measure up. That last bit is probably the hardest for any supervisor to do, but oh so necessary in preparation for the brutality of academia.

My second supervisor, Erez Levon, has similarly been an unfailingly attentive mentor in the four years I have been at Queen Mary. If it weren't for him, I would never have come to QMUL in the first place. Thanks for recruiting me, Erez! Thanks also for dedicating so much time, particularly in my first two years, for catching me up.

It is unlikely I would have pursued a Ph.D. – much less been equipped to pursue a Ph.D. – without the early mentorship of my M.A. supervisor Kari Fraurud. Kari, although we only worked together for a short time, you were the person who introduced me to sociolinguistics. Your research legacy planted a seed in me, and it has been growing ever since.

An important source of moral support has come from my colleagues in the Department of Linguistics. There were moments when I was gone for long periods of time, but they always made me feel welcome when I returned. I'd like give a big thanks to Agnieszka, Annette, Colleen, Christian, Danniella, David A, David H, Diana, Elisa, Elvis, Esther, Hazel, Jenny, Jonathan, Louis, Luisa, Maame, Margaret, Matthew, Mohammad, Pedro, Pietro, Reem, Rosie, Stamatina, Tom, and Ynda. Thank you guys for all the stimulating conversations and good company. Margaret, thank you for doing those last-minute edits when my brain was fried. I would like to especially thank Adam Chong for all of the insightful guidance regarding phonology and prosody.

I wouldn't had made it this far without my loving friends. Tharius, you're my Ace and my go-to for perspective on all things, great and small. I'll miss shading London with you! Li'l Nate, thanks for the regular visits and for your unwavering compass for reason. Habib, thank you for all the campy moments and for that MacBook Pro! André, thanks for being my sounding board and travel companion (and for putting up with my fussiness). Harish and Roger, our trash-TV binges kept my spirits up – as did the takeaway and champagne.

Dr. Brown and Dr. Ameneshoa, thank you for saving my life, and sorry for the occasional grouchiness. Roseanne, thank you for building my strength back up. Hugo, thanks for continuing what Roseanne started and for being such a great motivator.

To my Danish family, thank you for making sure I was never alone: Elsebeth, Søren, Inga, Niels, Marian, Alma, Sven, Lotte, and of course, lille Liv. All of you are blood to me. The visits and the phone calls and the care packages were a real source of motivation.

I'd like close by thanking my parents and my brother. Y'all have helped me in so many ways: mentally, financially, emotionally. For that I will always be thankful. When I fell ill, you dropped everything and nursed me back to health. So in that sense, you brought me to life *twice*. I think this dissertation is a great way to kickstart my second chance.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

This investigation was motivated by my growing sense that the sociolinguistic situation in Stockholm is both exceptional and exceptionally under-researched. Stockholm's sociolinguistic exceptionality owes primarily to its multiethnolect. 'Rinkeby Swedish' was Scandinavia's first and Europe's earliest-documented multiethnolect. Exceptionality has strong comparative connotations. Indeed, after 15 years of exposure to the multiethnic street vernaculars of Copenhagen, four years of exposure to Multicultural London English, and seven years of exposure to Stockholm's multiethnolect, my sense is that the latter is typologically more divergent from its mainstream counterparts than the former two.

This dissertation is on speech rhythm in one of Europe's most dynamically stratifying cities. The motivation for rhythm as a variable is both parochial and theoretical. As I will discuss in Chapter 5, 'staccato' rhythm is one of the most common objects of metalinguistic commentary in Stockholm. And beyond the narrow scope of the field site, it is generally accepted that the most important component to language change is *sound*:

> Readings in the work of 19th- and early 20th-century linguists make it abundantly clear that they saw sound change as the primary, most systematic and omnipresent mechanism of linguistic change. [...] [I]t can be argued that change in the surface phonetics remains the driving force behind a very large number of linguistic changes, perhaps the majority. This includes the processes of cliticization, which triggers any number of syntactic consequences, vowel contraction and consonantal assimilations, shifts of syllabicity and reassignment of syllable boundaries, along with the vast body of segmental changes – lenition and fortition, deletion and epenthesis, monophthongization and diphthongization, change of place and fusion of features, and the development of tone and its intersection with intonation patterns. (Labov 2001:11–12)

This argument is the other key motivation to the study at hand. Surface phonetic change will be foundational to any change, including the exceptional change in Stockholm.

The sense that Stockholm is exceptional has only intensified since I first set foot there in 2012. I recently researched foreign lexical matter in Danish and Swedish multiethnolectal hip hop (Young 2018a) and found that Swedish had more than twice the number of foreign lexical borrowings than Danish. Swedish also had more functional borrowings like pronouns and verbs while Danish had almost solely content loans. In an ongoing expansion of the comparison to British *drill* music, foreign loans in Swedish appear also to be more numerous and grammatically deeper. Since contact phonetics like rhythm are so underspecified, it might help to view these lexical loans as canaries in the coal mine that is contact-driven change.

Since multiethnolect is the starting point, the key focus of this study is the speech of Stockholm's racialized working class. The full scope of the study, however, is much larger. I have conducted a systematic analysis of rhythm in the speech of men in Stockholm from a stratified sample of social classes and ethnicities. This means that the study sheds light on social dialects beyond multiethnolect. These include white working-class speech, middle-class speech, and upper-class speech. In targeting the development and evolution of rhythm in multiethnolect, the analysis has necessarily investigated the rhythm of these other social dialects. They are just as important to the stratified linguistic ecology of Stockholm and are comparatively essential if one is to understand the rhythmic properties of multiethnolect.

I attempt to address four research questions that each pertain to speech rhythm in the context of interspeaker variation, intraspeaker variation, and change in apparent time.

> Rhythm in the vernacular: *Does the speech of Stockholm's racialized working class have lower rhythmic alternation than other varieties in the city; i.e., is it 'staccato'?*
>
> Phonology of Rhythm: *Is the variation in rhythm caused by (a) specific change(s) in the vowel system of certain speakers?*
>
> Rhythm in Contextual Style: *To what degree is speech rhythm stylistically sensitive in Stockholm, and what inference can be made about the maturity of its associated variants?*
>
> The Evolution of Rhythm: *What is the history of rhythmic stratification in Stockholm, and how has it evolved?*

In addressing the above questions, the final research output is a systematic account of rhythm in the speech of men in late-modern Stockholm. I unpack its social stratification, its variation across contextual styles, its phonetic components, and I attempt to trace its evolution by examining clues within age patterns and style shifting.

Roadmap

The dissertation begins with a focus on five characteristics of contemporary Stockholm that are uniquely late-modern. These are (1) large-scale migration and subsequent *superdiversity*

(Vertovec 2007), (2) spatial segregation and the residential enclosure of migrants, (3) school liberalization and subsequent educational hyper-segregation, (4) the emergence of a new elite, and (5) the emerging salience of whiteness within the traditional industrial working class. I argue in this first chapter that these five social facts have coalesced to make Stockholm rife with diverging socio-symbolic practices, one of which is speech practice.

In Chapter 3 I provide a review of the four social dialects that are said to circulate in Stockholm. These are Standard Swedish, traditional working-class *Ekensnack* or *Södersnack*, upper-class *Östermalm* or *Lidingö* Swedish, and the newest of the four, Stockholm's multiethnolect. The phonetic portion of the review is concentrated in the section on Standard Swedish, offering a 'crash course' in vowel and intonational phonology. This serves as an informational base for being able to follow the remainder of the dissertation. Only sparse accounts of phonetics and phonology are provided about Södersnack, Östermalm Swedish, and multiethnolect because very little phonetic research has been conducted on these varieties. Therefore, for these three social dialects, particularly multiethnolect, I focus mostly on language ideology.

Chapter 4 covers the data collection, transcription, and phonetic segmentation of the research corpus. The Chapter is broad and focuses on interview techniques, the development of a forced aligner for Swedish (SweFA, Young and McGarrah 2017), and addresses the omission of women's speech from the dataset. The chapter does not, however, cover question-specific methodologies. To facilitate easier reading, I have decided to allocate small Methods sections to each of the four analysis chapters.

Rhythm is a latent construction, and its study has not been without controversy. For this reason, the analysis portion of the dissertation is further delayed by a sizable literature review on approaches to investigating rhythm. Chapter 5 reviews isochrony theory, notions of compressibility, and rhythm metrics. This follows with a discussion of prominence and how prominence can be conceptualized. The chapter closes with a motivation for operationalizing prominence with Mishra, Sridhar, and Conkie's (2012) *energy-*f_0*-integral* while operationalizing rhythm with Low et al.'s (2000) well-known algorithm, the *normalized pairwise variability index for vowels* (nPVIV).

Chapters 6, 7, 8, and 9 constitute the four main analysis chapters. Each of the research questions that I listed above are addressed one by one in each of the chapters.

Chapter 6 addresses the question of rhythm and interspeaker variation. I conduct two analyses. In the first analysis, I use a battery of social predictors to test whether rhythm stratifies according to social class and the binary racialized schema, *svensk* and *invandrare* – two terms that I review in Chapter 4 from a Critical Race theoretical perspective. In the second analysis, I use social network data to cross-validate the findings in the first analysis. The results show that speech rhythm stratifies in three ways in Stockholm. It splits within the working class along racial lines, and elites produce a rhythmic pattern that lies between those two working classes.

Chapter 7 investigates the phonological transformations behind the dataset used in Chapter 6. Its key focus is on durational transformations. Four questions are explored: (1) are phonologically long vowels shortening in multiethnolect, (2) are phonologically short vowels lengthening in multiethnolect, (3) are phonologically long vowels more diphthongal in the traditional working-class variety, and (4) are phonologically long vowels more monopthongal in multiethnolect? I conclude that the questions are too simplistic and that the key transformation behind the change in rhythm in multiethnolect most likely lies in the shortening of prominent long vowels and the lengthening of non-prominent/unstressed short vowels. I find less evidence that prominent short vowels are lengthening or that non-prominent long vowels are shortening. Further, I find no evidence that either working-class variety is more diphthongal or monopthongal than the other. Two inadvertent discoveries, however, emerge in this Chapter. The first is that racialized speakers are leading in the Central Swedish lowering vowel shift. The second is that racialized speakers have a larger vowel space area (VSA), partly due to this lowering change shift and partly due to them producing fuller realizations of unstressed vowels (i.e., reducing them less).

Chapter 8 addresses the question of rhythm and intraspeaker variation. I conduct three analyses. In the first, I compare linear regression models between casual speech and the two reading styles to see whether social predictors have weakened. In the second analysis, I use a Random Forest variable importance analysis to test whether all social predictors attenuate in their predictive power of nPVIV as formality increases. In the third analysis, I test contextual style as an interaction effect in a linear regression model. All three analyses point to rhythm carrying significant stylistic sensitivity in Stockholm.

Chapter 9 is the final analysis chapter and addresses change in apparent time. Its two key conclusions are that staccato rhythm has moved from indicator to marker over time and that the transformation of speech rhythm by racialized speakers has not been a uniform process. For older speakers, the staccato effect is achieved by reducing accented vowels along the parameters of duration, f_0, and intensity. For younger speakers, these accented vowels are actually realized more fully. But their speech is overall much more 'staccato' due to a striking enlargement of unstressed vowels – short and long – along the parameters of duration, f_0, and intensity. The chapter ends with an exploratory analysis of the effects of network structure on the intermediate alternation pattern of higher-class *svensk* speakers.

Chapter 10 contains a summary of the dissertation findings. Chapter 11 contains a discussion of the following seven points: the methodological contributions of this dissertation, theoretical contributions to traditional Swedish dialectology, the epiphenomenality of rhythm, the theoretical implications of split findings within the working class, the tie between 'multiethnolect' and these findings, the role of *generation 2.5* in the focusing of multiethnolects, and the role of bourgeois political movements in creating new social dialects.

Chapter 2

LATE MODERNITY IN STOCKHOLM

Shortly after relocating to Stockholm in 2012, I happened across the heading 'Most people know what social class they belong to!' in the tabloid *Dagens Nyheter*. The article reports on the new dating site *mindAlike*, an exclusive service that requires its members to earn a minimum of 500 000 SEK per year and have an academic degree. It explains that this sort of class-based screening is common in other countries but relatively new to dating sites in Sweden. In that same year, I happened upon an advertisement in the metro that had been vandalized. The tutoring company *StudyBuddy* was running a new marketing campaign with the slogan 'dedicated, individual-tailored tutoring for reaching your goals!'[1]. The final phrase had been crossed out and replaced with new text: 'dedicated, custom-tailored tutoring for ~~reaching your goals!~~ a dystopian class society!'. Since the 1960s, municipal schools had offered free universal after-school tutoring for students struggling with homework. A recent law by the center-right coalition had eliminated this service and replaced it with a private voucher system.

In the following year, May 2013, youth riots broke out in Husby, the suburb in which I was residing. In reaction to the police murder of an elderly man and subsequent attempted cover-up, youths took to the street for five nights burning cars and throwing rocks. The events in Husby fascinated a global media that was otherwise accustomed to rosy stories about Sweden. The riots dominated headlines across the world for days. Shortly after, sociologists from the Universities of Edinburgh, Goldsmiths, Kings College, Royal Holloway, Gothenburg, London School of Economics, Malmö and Uppsala penned a joint editorial

[1] 'Engagerad, individanpassad läxhjälp mot ~~önskad målbild!~~ ett dystopiskt klassamhälle'

in *Svenska Dagbladet* in which they provided some structural context behind the 'suburban revolt'.

> Whereas the social and educational contributions once made by youth and community workers have been reduced, the police's presence and visibility have increased. What used to be tackled primarily as social problems is now considered to be issues of law enforcement and order. The French urban researcher Loïc Wacquant characterizes this as a political shift from Keynsian 'welfare' to neoliberalism's 'workfare', whereby the State develops a Janus face. On one side, liberalizations provide greater freedoms for the social groups who are winners amid the societal changes in recent decades. On the other side, [they provide] a reduced freedom for people in impoverished urban communities with reduced benefits and an emphasis on obligations linked to 'work first' in Sweden[i] (my translation, Back et al. 2013).

These media artifacts reflect a city and nation undergoing radical societal change. In the last thirty years, Sweden has transformed itself from a relatively homogeneous welfare state with a large manufacturing-based economy to a multiethnic, economically liberalized welfare state with booming creative and tech sectors. Much of this transformation started in the 1970s as Sweden began to cope with the emergence of globalization. The global shift to relocate industrial production in developing nations forced some of Sweden's largest industries to move their factories abroad or risk going bankrupt. Sweden responded by diversifying its economy and liberalizing the energy, transportation, healthcare, housing, and education sectors. Amid all of this, Sweden was also undergoing a demographic transformation due to non-Western labor migration followed by non-Western refugee migration.

Across Europe, in fact, one of the defining characteristics of late modernity is the racialization of its class hierarchies due to significant non-Western labor migration (Hesse 2007; Lee 2010; Lentin 2008; Lentin & Titley 2011).

Late modernity in Sweden is also defined by this. However, unlike other Western countries, Sweden has simultaneously maintained a relatively strong welfare state. A key component to this was migrating union participation from manufacturing over to services[2]. According to Hedin, Clark, Lundholm, and Malmberg (2012), "with broad middle-class stakes in its comprehensive welfare system, the Swedish welfare state has proven to be resilient to far-reaching neoliberal reforms" (2012:444). They cite Harvey's (2007) conclusion that "Sweden is an example of what might be called 'circumscribed neoliberalization,' and its generally superior social condition reflects that fact" (2007:15). The key here is that Sweden's liberalization is radical *in comparison to how Sweden used to be*. This is important to remember because its liberalizations will appear mild in comparison to countries like the United States or United Kingdom.

[2]Illustratively, 1995 marked *peak* union membership for Sweden at 86% (Fulton 2013). This complicates the typical understanding of unions as a fixture in post-war manufacturing and illuminates one unique aspect of late-modern Swedish industrial relations.

This neoliberalization of society, combined with automation, globalization, and large-scale migration, is part of the decentralization that characterizes late modernity as an era. Bauman (1992) prefers the term *post modernity* and argues that the tendency towards decentralization and fragmentation is such a radical shift away from the *modern* vision that we ought conceptualize the two as entirely separate epochs.

> The most conspicuous features of the postmodern condition: institutionalized pluralism, variety, contingency and ambivalence – have been all turned out by modern society in ever increasing volumes; yet they were seen as signs of failure rather than success, as evidence of the insufficiency of efforts so far, at a time when the institutions of modernity, faithfully replicated by the modern mentality, struggled for *universality*, *homogeneity*, *monotony* and *clarity*. The postmodern condition can be therefore described, on the one hand, as modernity emancipated from false consciousness; on the other, as a new type of social condition marked by the overt institutionalization of the characteristics which modernity – in its designs and managerial practices – set about to eliminate and, failing that, tried to conceal. (Bauman 1992:187–188)

Stockholm, with 2.2 million residents in its extended metropolitan region, is Sweden's and Scandinavia's largest city. As the cultural and economic hub of the Nordic countries, it has experienced the forces of late modernity at intensive levels for some time now. In one respect, late modernity in Stockholm has striking similarities with other developed urban areas in the UK, Germany or even the US. But in another respect, the socioeconomic factors of the late-modern era have manifestations and magnitudes that are unique to welfare-state Sweden. Three main factors are especially meaningful for the development of Stockholm's multiethnolect: (1) Large-scale immigration, particularly from non-Western countries, (2) spatial segregation and the residential enclosure of migrants, and (3) school liberalization and subsequent segregation. My argument is that these factors coincided particularly early, setting off the tipping point that incubated Stockholm's multiethnolect.

Two additional late-modern factors are important for contextualizing the speakers of multiethnolect within a greater social-class ecology: (4) the emergence of a new elite and (5) the emerging salience of whiteness among the traditional industrial working class.

1 Mass migration from non-Western countries

Migration has always been an especially relevant issue for Sweden. Few other nations in the world experienced anything like the mass exodus of Swedes between 1850 and 1939. In this period, approximately 1.5 million people left Sweden, over a million of whom resettled in the United States (Sweden 2015). Poverty was the primary reason for this out-migration—Sweden sustained consistent and consecutive crop failures in the mid 1800's. Moreover, American republic ideals were appealing alternatives to Sweden's rigid class hierarchy (Barton, 1994)[3].

This early experience with emigration may have conditioned Swedes to have a generally more sympathetic view toward immigration. According to the policy institute *Fores*, Swedes have consistently had more positive attitudes toward immigration than any other nation in Europe (Fores 2015). Similarly, in the 2012 European Social Survey, only 1.5% of Swedes answered "none at all" to the statement "Should Sweden should accept many or few immigrants from poor, non-European countries?" compared to 4% in Norway, 8.8% in Denmark, and 10.4% in Finland (ESS 2012).

Like many other Western European nations, Sweden experienced a boom in immigration after World War II, increasingly from non-Western countries. The earliest cohorts of immigrants were from Europe, hailing from Finland, Italy, Greece, and the former Yugoslavia. Starting in the late 1960s and extending into the 1980s, Turkish guest workers began to arrive. After the 1970s, immigration began to be increasingly dominated by asylum-seekers rather than guest workers.

> The rise of refugees began in the 1980s when Sweden saw some of its highest immigration from countries like Iran and Iraq, Lebanon, Syria, Turkey, Eritrea and Somalia, as well as South American countries with repressive regimes. Today, some 45 000 people with Chilean background reside in Sweden, following the refugee waves caused by Augusto Pinochet's dictatorship of Chile during 1973–1990 [...] and today Sweden is home to the third largest Chilean community outside of Chile, after Argentina and the US. In September 1980, Iraq launched an attack on Iran that marked the start of a bloody eight-year war between the two countries. From 1980 through 1989, nearly 7,000 people from Iraq and 27 000 from Iran received residence permits in Sweden as refugees according to the Geneva Convention. [...] The US-led invasion of Iraq, which started in 2003, led to yet another wave of Iraqis migrating to Sweden. [...] The 1990s brought massive immigration from former Yugoslavia during the ethnic cleansing wars with over 100,000 Bosnians being granted asylum in Sweden alongside 3,600 Kosovo Albanians. (Sweden 2015)

These migration patterns were not merely happenstance. They were part of deliberate public policy whose aim was to establish a *multicultural* welfare state (Borevi 2014). This contrasts significantly with neighboring Denmark and Norway where the political discourse on migration has often identified Sweden as *antithetical* to their values. According to Anna Gaarsdlev, European Correspondent for *Denmark's Radio*,

> We have never seen ourselves as an immigrant country the way you in Sweden have. And it means we have talked about migration and refugees in a much

[3] As an example, women gained the right to vote in 1921, 12 years *before* the voting requirements for land ownership and a minimum income were abandoned in 1933.

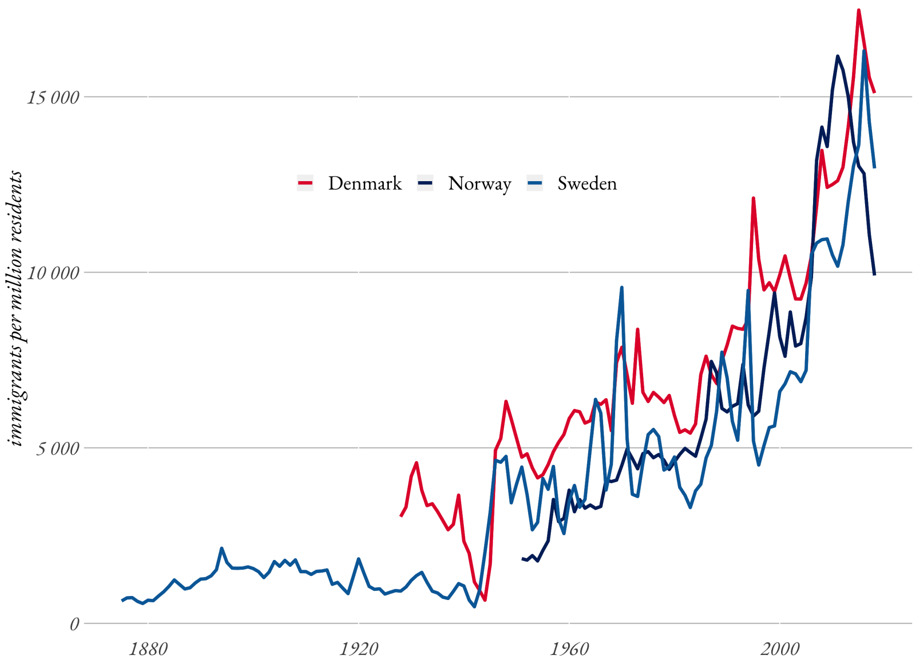


Figure 2.1: Annual immigration numbers for the Nordic countries, 1875–2018, normalized per million residents. Sources: *Statistics Denmark*, *Statistics Norway*, *Statistics Sweden*

> different way than you have. We have never felt that we have any special obligation here in the world to take in a lot of refugees. Danes simply do not feel that way. And that means that our perspective, or rather, the Danish people's perspective, to immigrants differs and has differed significantly from the Swedish perspective[ii]. (my translation, Forsberg 2014)

Many Swedes also see this as one of the key differences between them and their Nordic neighbors. Borevi (2014) notes that in contrast to Denmark where the merits of immigration were debated from the very beginning, there was broad political consensus in Sweden that immigration *and* the retention of one's pre-migration heritage was part of the modernist welfare model (2014:709–711).

This is to say that the differing attitudes towards migration are not due to different volumes of migration, as is often suggested. Figure 2.1 provides migration figures, normalized to population size, for each of the three Nordic countries. It shows similar rates between the three countries with Denmark having consistently higher numbers of incoming migrants per million residents. However, the limited data about *country of origin* paints a different picture. I mined historical data on country of origin for the residents of Copenhagen, Stockholm, Rinkeby and London in order to offer a comparative overview[4]. Figure 2.2 shows the num-

[4]In much of this chapter, I refer to other Nordic cities to compare Stockholm with its regional neighbors. I use London for supra-regional comparison; it is the first city of Europe while also happening to be Europe's most

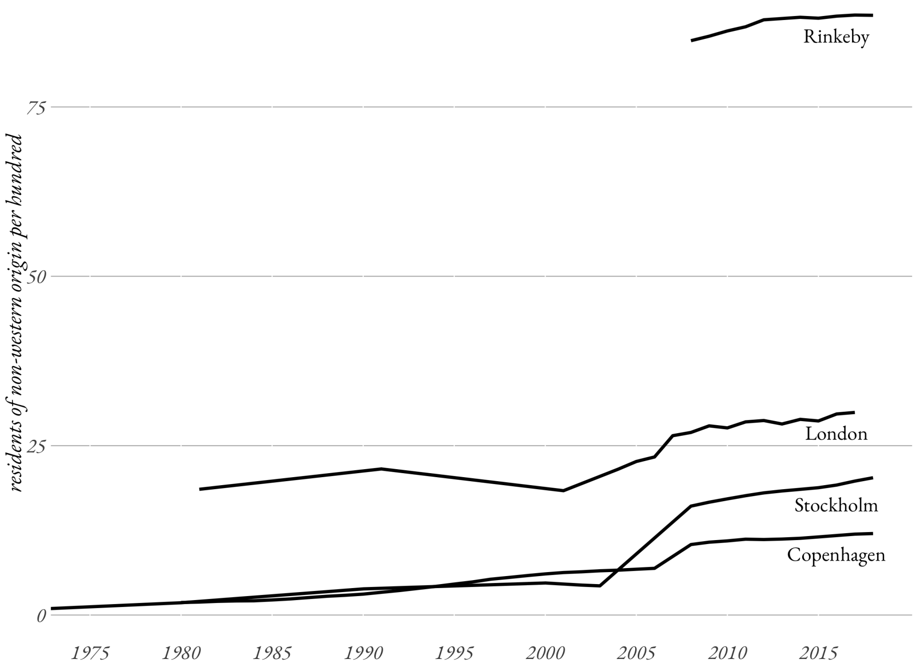


Figure 2.2: Residents of non-Western origin per 100 for Copenhagen (including Frederiksberg municipality), Stockholm, London, and the Stockholm neighborhood of Rinkeby, 1975–2018. Sources: *Statistics Denmark*, *Stockholms Stad*, and *UK Census*

ber of residents per hundred whose country of origin was not the European Union, European Economic Area, or North America. Recent data on the Stockholm suburb of Rinkeby is also provided, which I will discuss in Section 2 on residential segregation.

In 2018, 19.8% of the population in Stockholm municipality was of non-Western origin compared to the 12% in the Copenhagen/ Frederiksberg municipalities and compared to the 30% in London. Importantly, however, Copenhagen and Stockholm have had similar percentages of non-Western migrants up to 2004. This information contradicts the generally-accepted narrative that Stockholm has had relatively more non-Western migrants for longer than Copenhagen or Oslo. Quist (2012), for example, points to earlier waves of migration as one potential reason for the earlier development of Stockholm's multiethnolect.

> In Sweden this process started somewhat earlier, and one sees here a higher percentage of immigrants compared to Denmark and Norway. Immigrants and their descendants make up circa 11% out of 5.5 and 4.8 million people, respectively. In Sweden, the percentage is circa 17% of a total population of 9.2 million (numbers are from 2009/2013). These differences can possibly explain why it specifically was Sweden where the first investigations of youth language variation emerged[iii]. (my translation, Quist 2012:2)

multicultural city (Eurostat 2016).

Quist's 2009/2013 data is certainly correct; however, Sweden's higher migration figures do not appear to be caused by a head-start. Rather, the differences in 2009/2013 appear to be caused by an uptick in non-Western migration after 2004 in Sweden. A better explanation for the early inception of Swedish multiethnolect – and the comparatively *late* development of Danish and Norwegian multiethnolects – may have more to do with demographic tipping points. The Stockholm neighborhoods of Rinkeby and Flemingsberg[5] attained higher concentrations of non-Western migrants much earlier than the Copenhagen neighborhoods of inner Nørrebro, Kongens Enghave, and Avedøre and the Oslo neighborhood of Gamle Oslo.

2 The urban enclosure of migrants

Whereas the Copenhagen and Oslo multiethnolects first developed in historical working-class enclaves – *Inner Nørrebro*[6] and *Gamle Oslo*, respectively – Rinkeby Swedish is contemporaneous with the new establishment of Rinkeby and Flemingsberg. Built from the ground up in the late 1960s and early 1970s, Rinkeby and Flemingsberg attracted newly-arrived labor migrants alongside working-class transplants from across the city. On the other hand, in Nørrebro or Gamle Oslo, migrants and their children would have encountered long-fossilized working-class territories (and perhaps also the accompanying pressure to fit in linguistically). No pre-existing population existed in Rinkeby and Flemingsberg so there would have been more equal territorial footing between migrants and transplants from older neighborhoods. In that respect, the two suburbs have some similarities to the 'New Town' of Milton Keynes (Kerswill & Williams 2000)[7].

In addition to the power-leveling effect that universal newcomer status might have rendered, both Rinkeby and Flemingsberg are remarkably insular. Not only were they built along the city's periphery, they were designed in high-modernist fashion to be inward-facing and self-sufficient. Each suburb had its own library, day-cares, schools, medical and dentistry centers, recreational centers, and commercial district. Figure 2.3 shows a historical map of Rinkeby from 1971. What is immediately apparent about the suburb's design is that only *a single road* connects its 14 000 residents to the rest of the city. Not even neighboring Tensta is very accessible. Its other access point is via the Stockholm metro, which was connected to it in 1975. Aside from these two avenues, Rinkeby has remained a multiethnic island into the present day.

Between 1942 and 1977, over 65% of all housing in Sweden was built at a relatively large

[5]Despite what might be inferred from its name, half of the original speech data in the first study of Rinkeby Swedish was collected in the southern suburb of Flemingsberg (Kotsinas 1988a:266).

[6]Quist's (2000) seminal work on Copenhagen multiethnolect also relied on data from Kongens Enghave and the postwar housing estate of Avedøre Stationsby. However, these two neighborhoods had comparatively lower percentages of migrants than inner Nørrebro: 44% and 44.8%, respectively, versus 55.9% (Quist 2000:146)

[7]However, Milton Keynes was an actual new city with a full social-class hierarchy and mostly regional migrants. Rinkeby and Flemingsberg, on the other hand, were/are exclusively working-class enclaves.

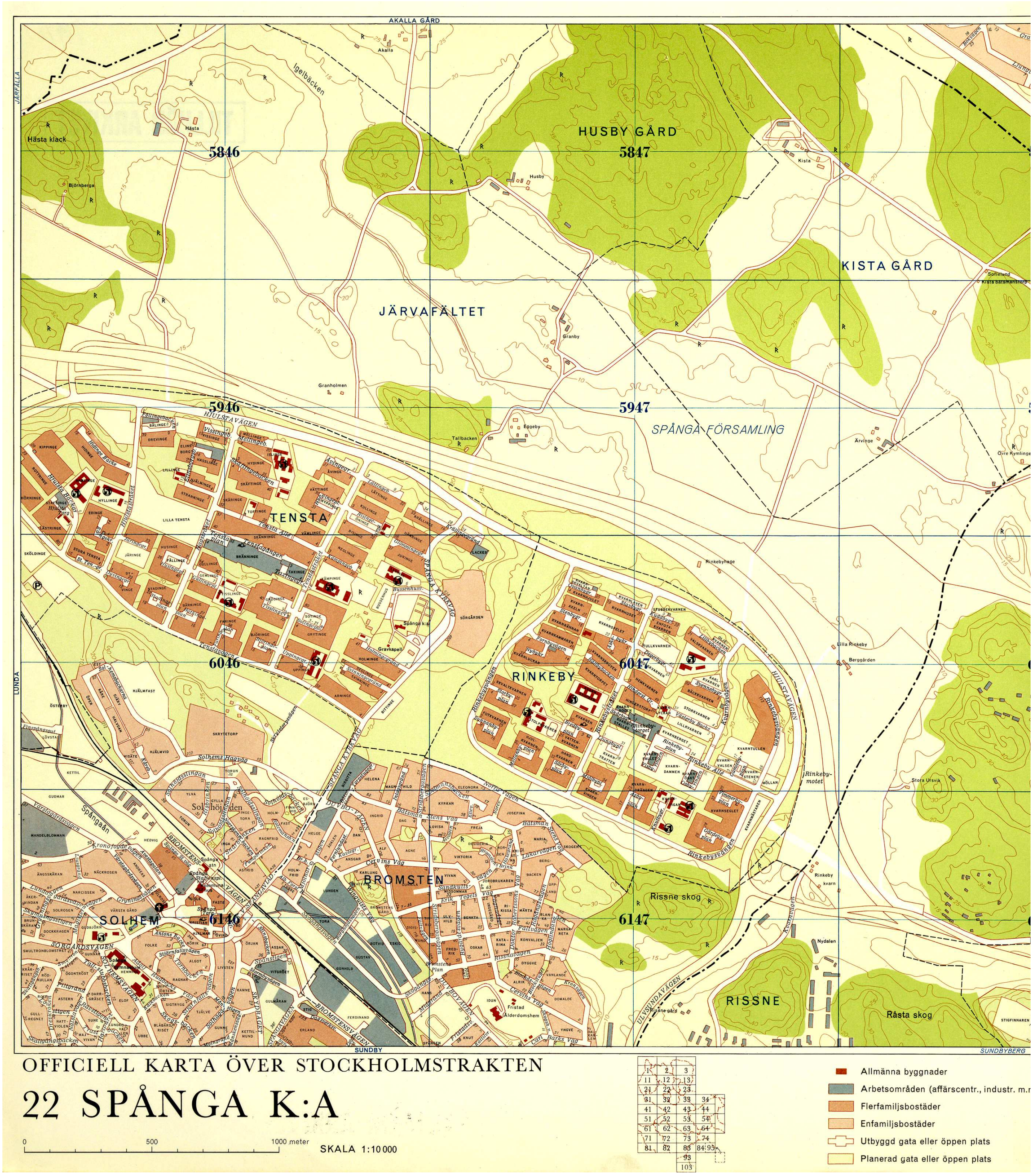


Figure 2.3: Map of the new suburbs, Rinkeby and Tensta, 1971 (Stadsmätningsavdelningen i Stockholm 1971)

scale by either public authorities or resident-owned cooperatives (Headey 1978). It was at the tail-end of this period that the famous *miljonprogram* (*Million Homes Program*) took place (Hall & Vidén 2005). Over one million housing units were constructed between 1965 and 1974 (Statistics Sweden, 2010:185) in what is considered to be one of the greatest achievements of Swedish socialism. Due to the refugee wave of the 1990s, many more of these suburbs came to be inhabited by immigrants and their descendants, which is why Swedish multiethnolect is sometimes colloquially referred to as *miljonsvenska* (Million Swedish). Most of these suburbs were laid out in the enclosure style of Rinkeby and have also become multiethnic enclaves[8].

A unique aspect of late modernity in Stockholm is that the concentration of migrants within peripheral housing projects has coincided with the most rapid increase in social inequality Sweden has ever seen. According to the OECD (2011), income inequality in Sweden grew at a breathtaking pace between 1985 and 2008. In 1985 the Gini coefficient[9] for Sweden was 0.20—the lowest in both Scandinavia and the world. By 2008, it had climbed to 0.26. This figure in isolation is still relatively low, but the 0.06 climb between 1985 and 2008 represents (tied with New Zealand) the highest change of any developed country in the report. By comparison, the UK's Gini coefficient grew by 0.02, the US by 0.04, Denmark and Norway by 0.03, and Finland by 0.05 (OECD 2011:24).

This growth in income polarization has coincided with residents relocating themselves on a massive scale to live closer to people who resemble themselves both racially[10] and socioeconomically. Hedin et al. (2012) investigated changes to the income distribution per geolocated hectare within Stockholm, Gothenburg, and Malmö across three five-year periods: 1986-1991, 1991-1996, and 1996-2001. They found increasing polarization in all three cities during this time. They refer to two processes at work behind this polarization: *gentrification* and *filtering*. Gentrification refers to an increase in mean income within a hectare during the five-year time period, and filtering means a decrease in mean income. *Super-gentrification* refers to an increase in income in an already high-income hectare, and *low income filtering* refers to a decrease in income in an already low-income hectare. The rate of filtering and gentrification in Stockholm increased significantly between 1986 and 2001, illustrated in Table 2.1. It shows that the number of socioeconomically stable neighborhoods decreased from 86.% in 1986 to 76.9% in 2001 with a significant acceleration of filtering neighborhoods between 1991 and 2001 (Hedin et al. 2012:452).

Interestingly, when one compares Hedin et al.'s (2012) findings with existing data on migrant settlement in Stockholm, the parallels are striking. The neighborhoods hit hardest by low-income filtering are mostly neighborhoods with large migrant populations. Figure 2.4 shows the progression of low income filtering by five-year period. Figure 2.5 shows a demo-

[8] With the exception of Assyrians in Södertälje, Greater Stockholm has no monoethnic immigrant enclaves like those typically seen in other European cities (Aktürk-Drake 2018).

[9] The Gini coefficient is a widely-used index for measuring income distribution among a nation's residents. A coefficient of 0 means that every resident earns the exact same. A coefficient of 1 means that all the income is owned by single individual.

[10] See section 7.1 (p. 86) for a discussion and problematization of the term *race*.

	Stockholm	
1986–1991		
Gentrifying neighborhoods	2,418	7.2%
Supergentrification	722	2.2%
Ordinary gentrification	1,090	3.3%
Classical gentrification	606	1.8%
Filtering neighborhoods	1,996	6.0%
Low-income filtering	789	2.4%
Stable neighborhoods	28,981	86.8%
Total	33,395	100.0%
1991–1996		
Gentrifying neighborhoods	2,550	7.6%
Supergentrification	1,483	4.4%
Ordinary gentrification	863	2.6%
Classical gentrification	204	0.6%
Filtering neighborhoods	3,928	11.8%
Low-income filtering	1,850	5.5%
Stable neighborhoods	26,917	80.6%
Total	33,395	100.0%
1996–2001		
Gentrifying neighborhoods	4,941	14.8%
Supergentrification	2,329	7.0%
Ordinary gentrification	2,056	6.2%
Classical gentrification	556	1.7%
Filtering neighborhoods	2,785	8.3%
Low-income filtering	957	2.9%
Stable neighborhoods	25,669	76.9%
Total	33,395	100.0%

Table 2.1: Number and percentage of gentrifying and filtering neighborhoods in Stockholm: 1986–1991, 1991–1996, and 1996–2001 (reproduced from Hedin et al. 2012:452)

graphic color map of Stockholm based on the percentage of residents either born abroad or with two parents born abroad, according to data I harvested from Statistics Sweden for the period 2010–2014. Notably, all of the regions highlighted in dark green (high percentage of migrant residents) in Figure 2.5 are also highlighted in purple (low-income filtering) in Figure 2.4.

Both figures illustrate a sharply-segregated city today that is continuing to polarize. Areas like Danderyd, Lidingö, Sollentuna, Nacka, Saltsjöbaden, Bromma, and Östermalm show super-gentrification in Figure 2.4, and they show under 20% of 'foreign background' residents in Figure 2.5. In sharp contrast, Tensta, Rinkeby, Skärholmen, Fittja, Norsborg, Alby, Flemingsberg and Fisksätra show low income filtering in Figure 2.4 and well over 60% 'foreign background' residents in Figure 2.5.

In the United States, the coalescence of white abandonment with economic divestment is common, and it is often referred to as *white flight* (see Crowder 2000). Bråmå (2006) found a similar phenomenon when she examined the social and spatial isolation in Stockholm by analyzing citizen registers between 1990 and 2000[11]. She found that Swedes, defined as any individual with one or both parents born in Sweden, actively avoided neighborhoods with

[11] In Sweden, all citizens are registered centrally every time they move via *folkbokning* (citizen registry), and this registry is tied to your personal ID number to which demographic data is available to statisticians. National participation is mandatory and therefore nearly 100% of the populace.

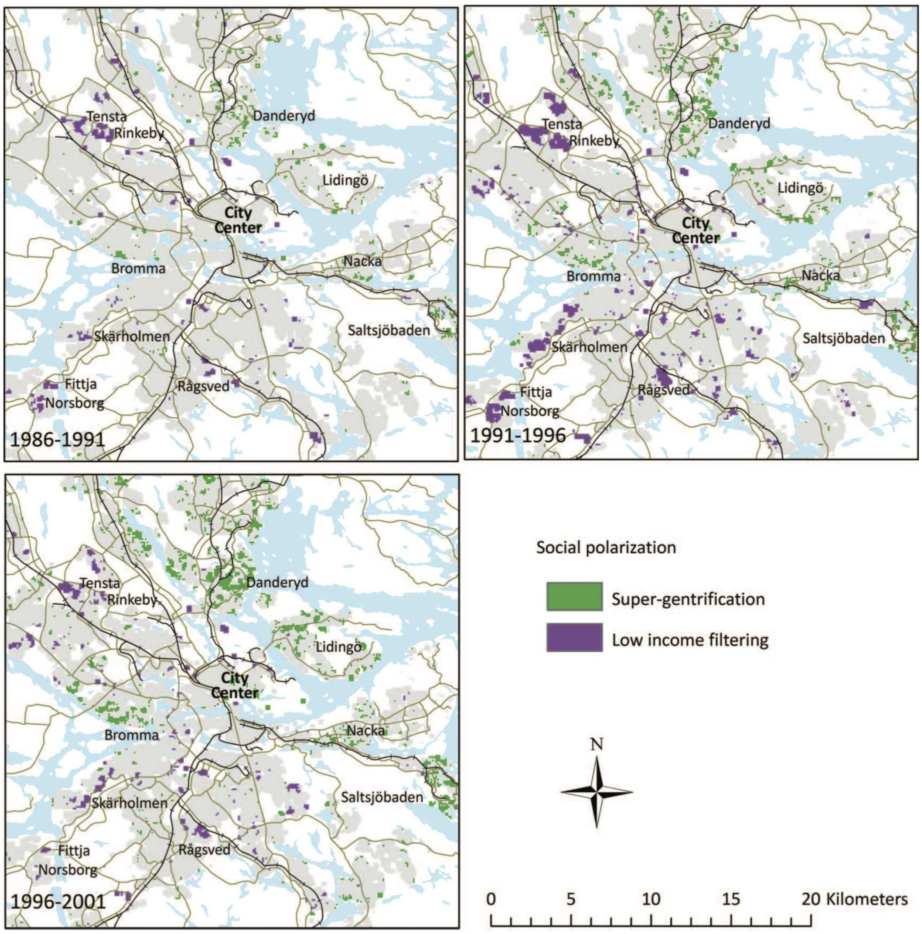


Figure 2.4: *Super-gentrification* and *low income filtering* measured by change in income per hectare in greater Stockholm between 1986 and 2001 (reproduced from Hedin et al. 2012:455)

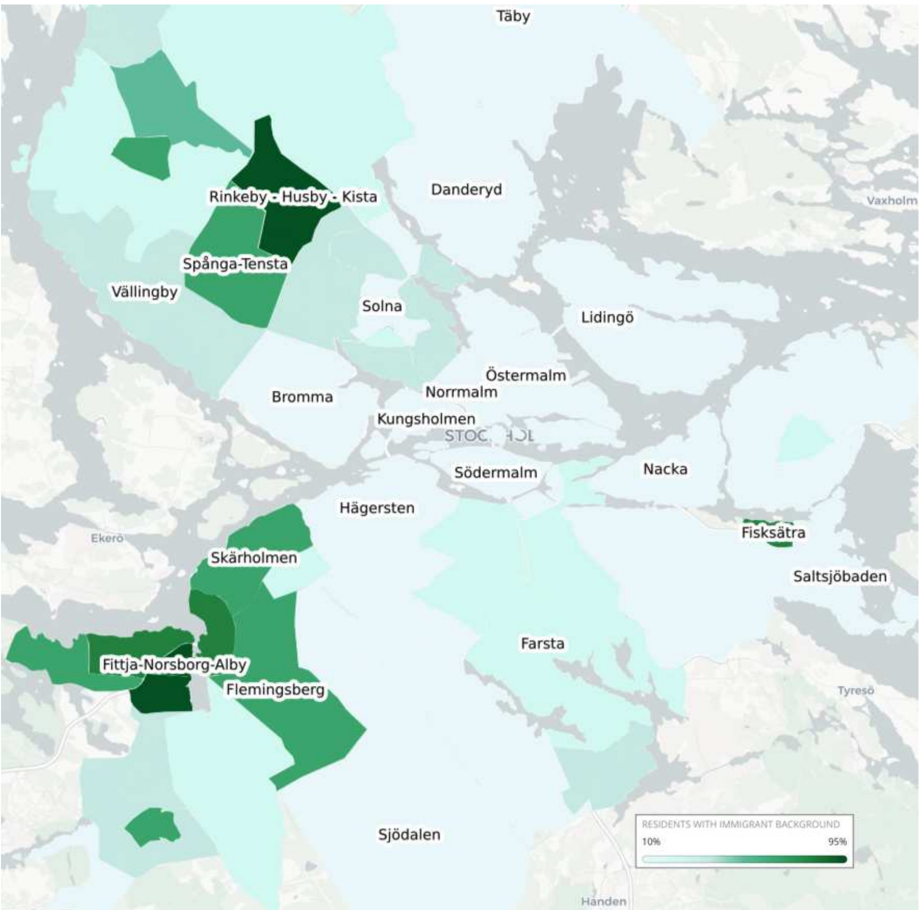


Figure 2.5: Percentage of residents either born abroad or with both parents born abroad by neighborhood in metropolitan Stockholm for the period 2010–2014

large immigrant populations. She called this phenomenon *Swedish flight* or 'Swedish avoidance'. She highlighted Husby – where the Stockholm Riots started in 2013 (p. 2) – as an emblematic example.

> In 1990, residents with a Swedish back-ground still were in the majority (53 percent), but this changed in the following year [1991]. In 1995, Swedes made up 33 per cent of the total population and, 5 years later, only 24 per cent. Husby was already multi-ethnic in 1990, with nearly 70 nationalities represented, and it remained so during the 1990s. The largest immigrant groups are those born in Iran, Iraq, Ethiopia, Somalia and Finland. (Bråmå 2006:1140–1142)

Her data indicated that 200 to 400 'Swedes' moved out of Husby every year between 1990 and 2000 and that 800 to 1300 Swedes of 'immigrant descent' moved in per year during the same period (2006:1142).

As I have documented in this section, a diverse range of sources identify inequality and segregation as key components to the late-modern era in Stockholm. The data have also documented that this period was unique in terms of its transformative intensity.

3 The liberalization and segregation of schools

The Swedish education reform of the 1990s is not unlike what occurred in the United Kingdom during this same period (see Rampton 2006:2–38 for a thorough review). Before 1992, a law by the name of *Lex Pysslingen* mandated that schools be run centrally and exclusively by the state. The statute was abolished in 1992, opening the doors for profit-driven schools to receive state vouchers for every student they enrolled. The national deregulation still allowed municipalities to regulate what extent they would participate in the subventionalized school model. In 2000, however, Stockholm removed all municipal restrictions, and in 2006 Parliament abolished the right for municipalities to restrict the voucher program in entirety. By 2008, nearly half of the municipalities with the highest percentage of subventionalized *friskolor* (free schools) were in the Stockholm region. For secondary schools, those municipalities were Täby at 63%, Danderyd at 47%, Sundbyberg at 43%, Nacka at 42%, and Stockholm at 40% (refer to the maps in Figure 2.5 for a bearing on these neighborhoods).

Numerous studies have shown that education reform in Sweden has contributed to (and still is contributing to) ethnic and socioeconomic segregation in schools. According to reports by the Institute for the Evaluation of Labour Market and Education Policy (IFAU), based on data between 1993 and 2009, a strong correlation was found between high *friskola* enrollment and increased school segregation (Böhlmark, Holmlund, & Lindahl 2015; Holmlund et al. 2014). In a separate report, the OECD (2012) cautioned that inequality in schooling had rapidly grown in Sweden since a decade before. It cited a number of sources linking the school reform at the end of the twentieth century with growing ethnic segregation. One

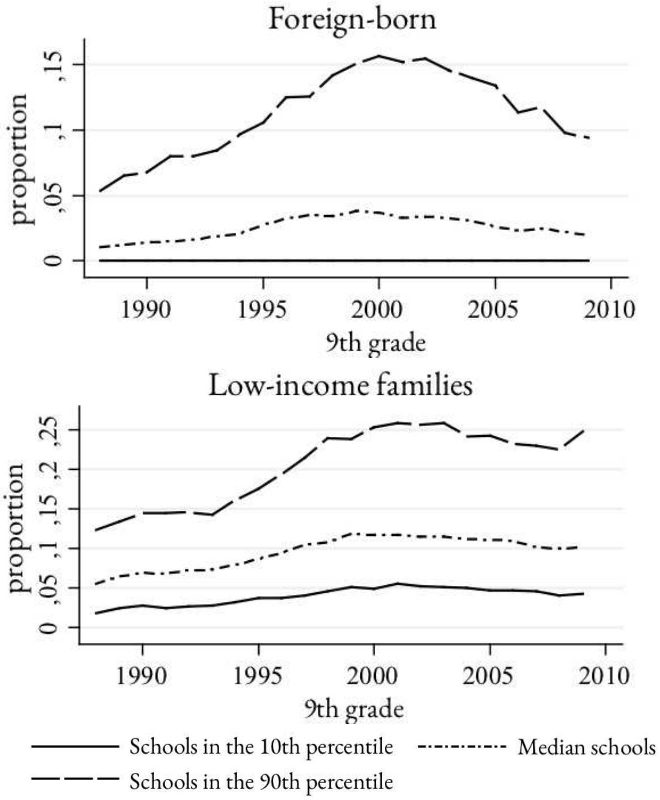


Figure 2.6: School segregation metrics for all primary schools in Sweden, 1988–2009 (reproduced and translated from Holmlund et al. 2014:88–90). *x*-axis is the year of the measurement. *y*-axis is the proportion of students in each school type as measured in the ninth grade.

of the sources it cited, Söderström and Uusitalo (2010), examined data on schools in Stockholm in 1998, 1999, 2000, and 2001 and found compelling evidence that the 2000 municipal deregulation had a direct effect on ethnic segregation in schools.

Both Holmlund et al. (2014) and Böhlmark et al. (2015) found clear turning points at approximately year 2000 for school segregation. Figure 2.6 contains a reproduction of two graphs from Holmlund et al.'s (2014) report and shows a peak at years 2000–2004 for segregation in terms of ethnic and socioeconomic lines. The top chart shows that the percentage of foreign-born students in the worst-performing schools (90th percentile) across Sweden climbed from 5% in 1988 to over 15% in 2000. The bottom chart shows that the percentage of low-income families in the worst-performing schools (90th percentile) across Sweden climbed from 12% in 1988 to 25% in 2000.

Forsberg (2018) harvested detailed demographic data from all of Stockholm's secondary-school pupils between 2006 and 2008 and found a direct link between *friskolor*, socioeconomic polarization, and gender segregation. Traditional elite schools were unaffected by the reform and maintained their status. They continued to draw in students from the most wealthy families. Many of the for-profit charter schools, however, under pressure by shareholders to fill their rosters, marketed themselves aggressively in low-income neighborhoods. This led to an abandonment of municipal comprehensive schools by the lowest social classes in favor of smaller, customized schools that are both class- and gender-segregated (Forsberg 2018).

Chapter 9 identifies some generational differences in multiethnolectal prosody that appear to coincide with some of the educational change points identified in this review. As I show in the next section on elites and the rise in their disposable income, the year 2000 appears to be a particularly transformative year in Swedish social history.

4 A new elite challenges the old order

Within the last two decades, young elites have taken a more conspicuous seat in Stockholm's collective consciousness. While typically the object of scorn, referred to at times as *Brats* or *stekare* (show-offs), they are the economic winners in a newly liberalized Sweden. The segregation discussed in the prior sections has also been accompanied by a massive private equity and technology boom in Stockholm. This has set in motion the establishment of a salariat with a purchasing power only previously held by magnates and aristocrats. According to Gustafsson and Jansson's (2008) longitudinal analysis of income micro data in Sweden,

> since the first half of the 1980s real income at the top of the distribution has developed more favourably than for other groups. This contrasts with the changes which occurred prior to the 1980s. Reasons for the rise in the top income share are several: the development of stock prices, tax reforms and the labour market change of top wages increasing more rapidly than others (Gustafsson & Jansson 2008:963).

Figure 2.7 is reproduced from Gustafsson and Jansson (2008:966) and shows the percentage distribution of disposable income for the two lowest and the two highest deciles of the population in Sweden from 1975 to 2004. Starting in 1990, real disposable income began to decline for the bottom two percentiles. Around that same time, the slope of increase for the top percentile began to steepen, peaking sharply between 2000 and 2001. The authors observe that

> this indication has parallels in results from sociological research on people's attitudes to the welfare state. Results from surveys often indicate that the income of the respondent is positively related to acceptance of large income differences in society. In addition, Svallfors (1998) reports that the small number of respondents belonging to the elite group in surveys made for various years, have begun to have a more negative attitude to collective funding and to harbour more suspicions of abuse of benefits and services than other respondents. (Gustafsson & Jansson 2008:967)

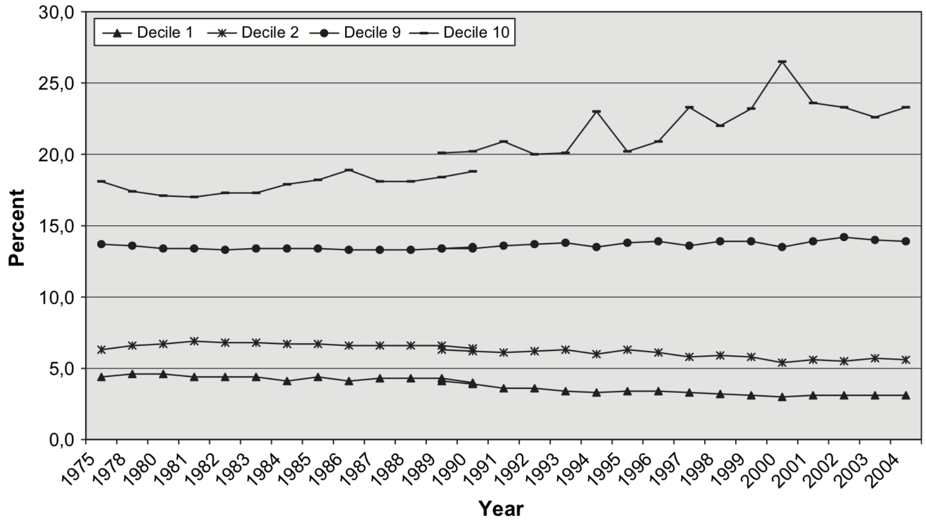


Figure 2.7: Income share (disposable income) for deciles 1, 2, 9 and 10 (1975–2004), reproduced from (Gustafsson & Jansson 2008:966)

This shift in values among some of Sweden's elites is also accompanied by ostentatious behavior that otherwise would have been taboo in an earlier time. Ostberg (2007) observes in his ethnographic investigation of *Brat* subculture that the conspicuous consumption of Stockholm's new elite reflects a rebellion against the old and 'outdated' egalitarian system that their parents embraced.

> The Brats' inclination toward conspicuousness should be read against this relative downplaying of the style leadership of their parents. [...] This can also be seen in light of the larger backdrop of Sweden's longstanding political tradition of relative economic and gender equality. The Brats are, more than anything else, rebelling against this ethos of equality and standing up for their right to conspicuously flaunt their stuff; a right that their parents swindled them of. The worst accusation for the Brats is to be accused of being just one of 'the Jones', or 'A Svensson' as they would have it in Sweden (Ostberg 2007:100)

Performances of wealth are concentrated in specific parts of town, the most famous of which is the central neighborhood of Östermalm and its iconic entertainment district *Stureplan*. Other areas include the suburbs of Bromma, Danderyd, Lidingö and Saltsjöbaden. Referring back to Hedin et al.'s (2012) investigation of super-gentrification in Stockholm, these are exactly the areas that have experienced the most extreme growth in mean income (marked in green on Figure 2.4, p. 28).

Matory (2015) expresses surprise that the wealthy are so often a conversational topic in Stockholm. In contrast to what he predicted to be a focused derision toward migrants, he encountered repeated antipathy toward so-called *Stureplan people* instead.

> During last September's national elections, while the Swedish news media focused on the refugee question, the main social antitype I heard about as I circulated through the restaurants, bars, nightclubs, and shops of both the elite Östermalm and the artsy-but-gentrifying Södermalm neighborhoods were the Stureplan people. They give a face to the 'matter out of place' that is the hyper-conspicuous, moneyed elite — a group that is not supposed to exist in egalitarian Swedish society. (Matory 2015:51).

Elites in Stockholm, while overwhelmingly of Swedish ethnicity, are becoming more and more diverse across ethnic lines. In fact, Crul, Keskiner, and Lelie (2017) found that the children of Turkish migrants in Sweden were much more likely to work in managerial positions than in the rest of Europe. In Sweden, 24% were employed within the top two tiers of the EGP (Erikson-Goldthorpe-Portocarero) Occupational Class Coding scheme. In contrast, the figures for France, Germany, and the Netherlands were 18%, 12% and 19%, respectively. The numbers are even higher for Swedes of Iranian descent – a group viewed as emblematic of Sweden's immigrant elite. In 2010, 54% of Swedes of Iranian heritage were enrolled in higher education by the age of 25 compared to 43% for all ethnicities and 44% for ethnic Swedes. (Statistics Sweden, 2012). These graduates go into fields like medicine, banking, and private equity and have spent enough time abroad to know that their Swedish nationality, sharp fashion sense, and stellar credentials give them 'super white' status in places like New York and Singapore (Lundström 2010).

This group receives a fair amount of disdain from working-class Swedes of migrant origin. In my earlier work (Young 2014), the participants debated whether Iranians were 'real blattar' (Pakis) in the endearing, self-appropriating sense of the word. A short excerpt of the exchange is provided below. Amina and Reman are brother and sister and are of Kurdish heritage. Hayder is of Kurdish heritage, and Murad is of Egyptian-Arab heritage. Saman is Reman's girlfriend and is of Iranian descent. All were born in Sweden.

NATE: Why are you asking whether an Iranian is a blatte? That is interesting.

AMINA: Because they are much more assimilated than anyone else from the Middle East.

HAYDER: Yup. They are the most established.

REMAN: I mean, they forget their roots as soon as they...

AMINA: Assimilated is, well...

HAYDER: I mean, they are...

MURAD: They are better than us on all fronts!

Lots of laughter

MURAD: (during laughter) I mean, they are...

Continued laughter

MURAD: (during laughter) Their country is way back there, and totally zero compared to Sweden!

AMINA: (directed to Saman in stylized other voice) You're all so civilized!

SAMAN: (directed to Amina in stylized other voice) You're all animals!

HAYDER: It's been that way; long before the Ayatollahs (referring to the Islamic revolution)[iv].

(Young 2014:45)

The last line refers to the fact that many of Iran's intellectual elites migrated to Sweden after the Islamic Revolution, which provides a likely explanation for the upward mobility of its second generation. As the above exchange shows, it has resulted in the situational negotiation of racialized boundaries, since non-whiteness is so often seen to be the same as local and working class. This theme of exceptional Iranians is also alluded to in popular culture like this excerpt of a recent hip hop song:

Two Chileans from the Red Line[12], yeah they were cousins
They used to import coke as if it were oranges
Everyone got involved; the whole family cashed in
But now they got caught; grandma's riding in the paddy wagon

Two Yugoslavs live in a mansion now, Soran and Goran
They were benching 150 without anabolics
They've been around since the 80s; used to fuck Carola
They made money before anyone else had any

Arash from Hallonbergen was a Svenne-Iranian[13]
He's been talking circles around Swedes since middle school
Was the best salesman in his company and made five million
Now he lives on Strandvägen[14] with sharks in his aquarium[v]

(STOR 2016:1:29)

What the above discourse also alludes to is the problematic *husblatte* construction ('house Paki'). A calque of the American slur *house n****r*, parallel terms in the UK and Denmark are *coconut* and *præmieperker* (prize Paki), respectively. The term has been used in the suburbs for quite some time but only came to the attention of mainstream Swedish society in 2015 when Kurdish-Swedish activist and comedian Özz Nûjen publicly called the political pundit Alice Teodorescu a *husblatte* (Andersson 2015). He argued that her conservative views made her a puppet for white Swedes and a traitor to her racial, and *prima facie* working-class, affiliation. An intense denouncement followed, particularly from the Center-Right establishment.

Authenticity among second-generation migrants is tightly bound up in working-class identity in late-modern Stockholm. This, however, is at tension with the reality that many immigrant groups, such as the Iranian and Chilean intelligenstia, were already better equipped to deal with Sweden's neoliberal reforms than migrant groups from more humble class backgrounds. What is more, much of the migrant intelligentsia lost class privilege when they

[12] *Red Line* is a reference to the southern stretch of Stockholm's red metro line T13 that travels through the southern immigrant suburbs (Bredäng through Norsberg, refer to the metro map in Figure 2.8).

[13] Approximate English translation for *svenne* is *cracker* or *honkey* and is used to designate a Swede who has either an especially light compexion or is especially stiff in body stylization or both.

[14] Strandvägen is one of the most elite addresses in Stockholm.

relocated to Sweden in the 1970s and 80s, which may have motivated them to push their children up the social ladder into adulthood. Today, these young urban professionals are equally vigorous actors in the construction of the new Swedish glitterati.

5 The emerging salience of whiteness within the established working class

As Sweden's population continues to diversify, and a *visible* minority population grows in presence, whiteness has become a more and more important construction for the disenfranchised 'native' working class. The arrival of a phenotypically 'marked' population – in other words, visible minorities – has coincided with de-industrialization and a destabilization of working-class earnings. Recall Figure 2.7 and the visible decline of disposable income for deciles 1 and 2. These social facts have invigorated a new construction of Swedishness and whiteness that derives meaning from its juxtaposition to The Other. Watt, Millington, and Huq (2014:143) explore this construction of whiteness among London's Cockney diaspora in Essex and observe that "whiteness and its attendant racisms are formed out of complex ontological insecurities regarding felt losses—of community and nation—under globalization's juggernaut".

Stockholm's traditional industrial working class has undergone a displacement that resembles the Cockney march into Essex. Stockholm has long been a segregated city, and the southern borough of *Södermalm* was the home of its industrial working class. To offer an illustrative anecdote, one of this study's upper-class participants, 'Johan', said that when he played Monopoly with his grandmother as a child, she refused to buy property on any Södermalm streets – even if it meant forfeiting the game. Starting, however, with the Million Houses Program in the 1960s, and accelerated by gentrification on Södermalm, this community has become more and more displaced. After over 20 years of gentrification, Stockholm's cultural and media elites constitute the most visible population in Södermalm today. Only small pockets of the white working class remain in sundry rent-stabilized units[15].

Very little research – whether sociological or sociolinguistic – has been conducted on this population. But according to popular knowledge, their migration has targeted the suburbs to the south and southeast. Two transport corridors have facilitated this movement; the first is the southern stretch of Stockholm's green metro line, particularly the T17 to Farsta strand and the T18 to Skarpnäck. This is shown in Figure 2.8. The second corridor comprises of the two commuter train lines (pendeltåg) that run south of Stockholm (not shown in Figure 2.8). One splits west and runs through Tumba. The other splits east, running through

[15] To offer an anecdote, I lived for two years in a sublet on Södermalm near the Zinkensdamm station. The building was part of a cluster of postwar housing projects, and almost all of my neighbors were elderly and spoke the traditional working-class Södermalm variety. All of the surrounding pre-war buildings, however, had been renovated and resettled by elites. The surrounding businesses were nearly all exclusive in nature, including the well-known restaurant *Pane e Vino* and the upscale second-hand boutique *Beyond Retro*.

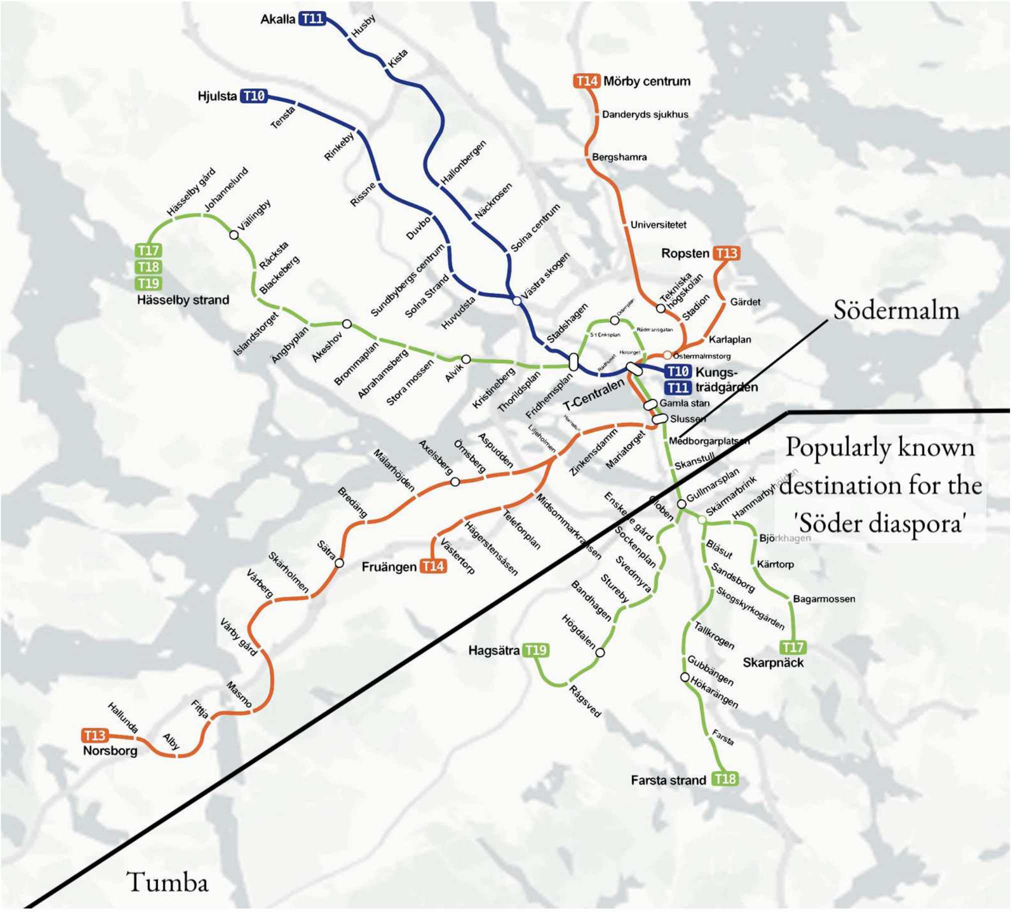


Figure 2.8: Migration pattern of the Söder diaspora – i.e., Stockholm's traditional industrial working class – according to popular knowledge.

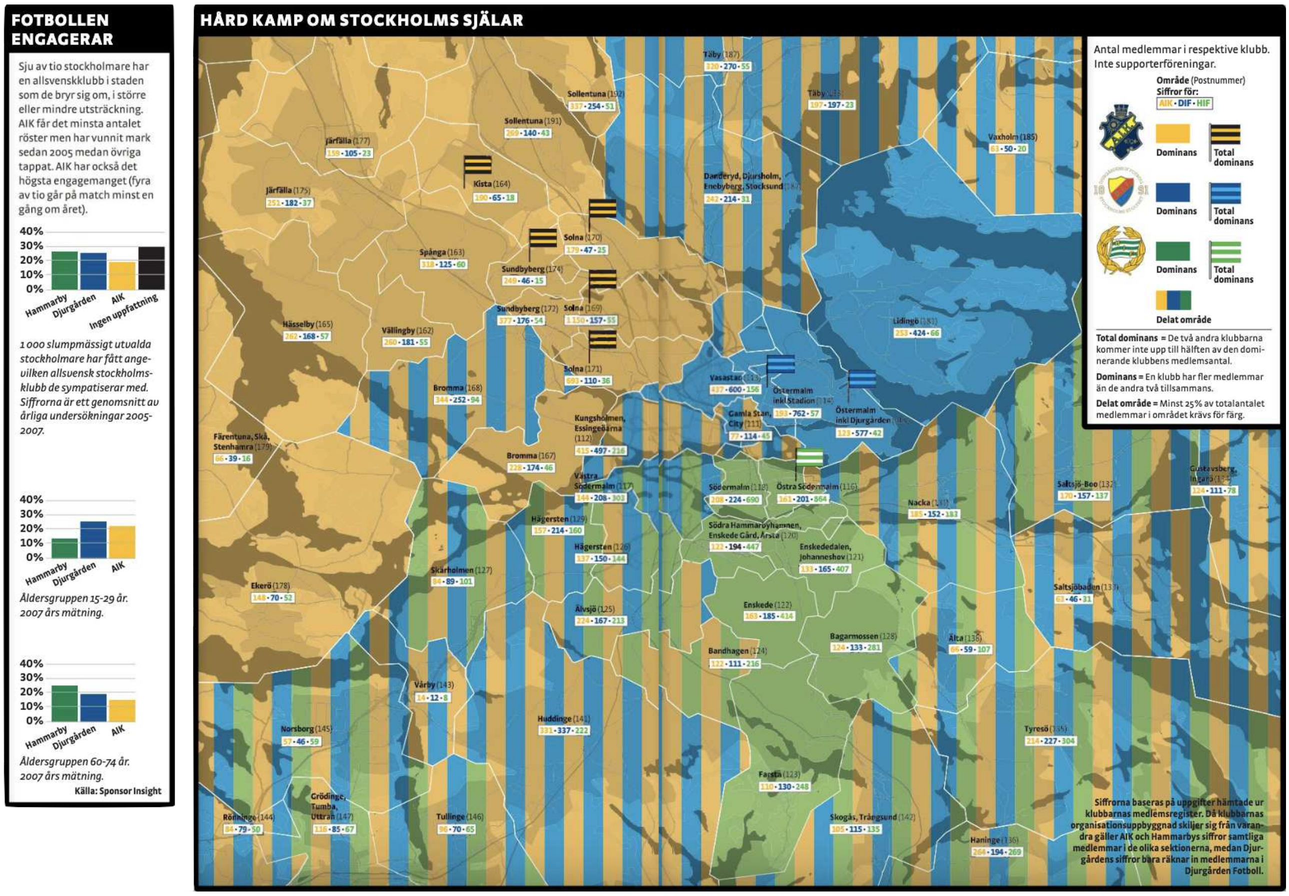


Figure 2.9: Heat map for fans of Stockholm's three main football clubs: *Hammarby* (green), *Djurgården* (blue), and *AIK* (*Allmänna Idrottsklubben*, yellow). The Södermalm-based Hammarby club has its densest fanbase in the southern suburbs (*Hård Kamp om Stockholms Själar* [Hard Fight for Stockholm's Souls], reproduced from Ystén 2008:74–75).

Farsta strand and terminates at Nynäshamn. In my ethnographic work, these areas were consistently identified by participants as places where white working-class people resided.

In striking similarity to the strong tie between Essex and West Ham United, the same attachment exists within the Söder diaspora for the Södermalm-based *Hammarby* football club. According to the popular football magazine *Offside*, these so-called *Bayen* fans are overwhelmingly concentrated in Stockholm's southern suburbs (Ystén 2008:74–75). Figure 2.9 shows a heat map of fan concentrations for the three main football clubs in Stockholm – Hammarby, Djurgården, and AIK (Allmänna Idrottsklubben). The areas with the densest Hammarby fanbase, marked in green, all lie within the southern suburbs (e.g., Enskede, Bandhagen, Bagarmossen, and Farsta).

These also happen to be areas with relatively high support for the neo-fascist Sweden Democrat party (*Sverigedemokraterna*; refer to Mulinari and Neergaard 2014). Open Society Foundations (2014) conducted a recent series of ethnographic studies on white working-class communities in six European cities: Aarhus, Amsterdam, Berlin, Lyon, Manchester, and Stockholm. In Stockholm they focused on Storvreten in Tumba, a neighborhood that has always been a stronghold for the Söder diaspora but in recent years has been rapidly diversifying.

> Storvreten in Tumba is the most heterogeneous area in Southern Botkyrka, and all the different types of housing that exist in urban Sweden are there. It is also the place where open conflicts between majority and non-majority Swedes are most noticeable, according to one interviewee, and it is also the strongest constituency for the Sweden Democrats: "*It is almost like a war between whites and non-whites in the public space*", Interview with Björn Lagerstedt, member of the Moderate Party, Tumba, 18 March 2013
>
> "*Bremora is a white neighbourhood with only Swedes. Now darkies have started to arrive. There are some who just have curtains on their windows, and just keep to themselves. It's getting insecure. This creates problems and it will grow in Southern Botkyrka*", Comment from a participant in Focus Group 6: Policing and security, 18 April 2013 (Open Society Foundations 2014:58–59)

If one returns to the map of residents of foreign descent in Figure 2.5 (p. 28), what also stands out is that the white working-class neighborhoods to the south are in darker shades of green than the affluent neighborhoods to the north and east. Observe specifically Farsta and the light green areas south of Fittja-Norsborg-Alby. While middle-class and upper-class neighborhoods in the North have stayed relatively homogeneous, this has not been the case for white working class enclaves such as those studied in Open Society Foundations (2014). Their neighborhoods are diversifying at rapid rates, which implies that they have a closer proximity to multiethnolect than any other white group. Likewise, it implies that speakers of multiethnolect will have more contact with the established working-class variety than any other variety in the city.

—

I have identified three key events in this review that I believe were pivotal to the establishment of Stockholm's multiethnolect (aka, *Rinkeby Swedish*). In doing so, a picture has begun to emerge about multiethnolect and its associated speech community. The picture, however, is incomplete and will be further developed in the next chapter on social dialects. I focus there on the local metapragmatic discourses that have circulated on semilingualism, Rinkeby Swedish, and multiethnolect. I then also review the brief literature on its specific features.

I have dedicated additional space in this chapter to identify two social groups that I consider to be in juxtaposition to the multiethnolectal speech community. These are the white working class and the upper class, and both are important reference groups for any analysis that seeks to offer a meaningful account of linguistic features in multiethnolect. In the next chapter, I will also provide a review of their respective social dialects.

Citations in original language

[i]Samtidigt som de sociala och pedagogiska insatser som tidigare gjordes i dessa områden av fritidsledare och fältassistenter reducerats har polisens närvaro och synlighet ökat. Vad som tidigare i huvudsak angreps som sociala problem betraktas nu som polisiära ordningsproblem. Den franske urbanforskaren Loìc Wacquant karakateriserar detta som ett politiskt skifte från keynesianismens 'welfare' till nyliberalismens 'workfare', där staten utvecklar ett Janusansikte. å ena sidan liberaliseringar som ger större frihetsutrymmen för de sociala skikt som är vinnarna i de senaste decenniernas samhällsutveckling, människor som bor i välmående (men alltmer isolerade och säkerhetsfixerade) stadsdelar. å andra sidan ett minskat frihetsutrymme för människorna i fattiga stadsdelar, med sänkta bidragsnivåer och betoning av skyldigheter, i Sverige förknippat med 'arbetslinjen' (Back et al. 2013).

[ii]Vi har aldrig sett oss själva som ett invandringsland, på samma sätt som ni i Sverige har gjort. Och det betyder att vi har pratat om migration och flyktingar på ett helt annat sätt än ni har gjort. Vi har aldrig tyckt att vi har en särskild plikt här i världen att ta emot många flyktingar. Det tycker inte danskarna. Och det gör att vår inställning, eller danska befolkningens inställning, till flyktingar skiljer sig, och har skiljt sig avsevärt från den svenska inställningen. (Forsberg 2014)

[iii]I Sverige er denne proces begyndt lidt tidligere, og man har her en højere andel af indvandrere sammenlignet med Danmark og Norge. Indvandrere og efterkommere udgør i Danmark og Norge cirka 11% af de samlede befolkningsantal på henholdsvis cirka 5,5 og 4,8 millioner mennesker. I Sverige er andelen cirka 17% af den samlede befolkning på cirka 9,2 millioner (tallene er fra 2009/2013). Disse forskelle kan muligvis forklare at det netop var i Sverige de første undersøgelser af unges nye måder at variere sprog på blev beskrevet. (Quist 2012:2)

[iv]Nate: Varför frågar du om en iranier är blatte? Det är intressant.

Amina: För att dom är mycket mer assimilerade än alla andra från Mellanösten

Hayder: Aa, dom är mest etablerade.

Reman: Asså dom glömmer sitt ursprung asså så snabbt dom...

Amina: Assimilerade e ju en (unclear)

Hayder: Asså dom e så...

Murad: Dom är bättre än oss hela vägen (unclear)!

Lots of laughter

Murad: (during laughter) asså dom...

Continued laughter

Murad: (during laughter) Deras land där ligger där borta, noll med Sverige!

Continued laughter

Amina: (directed to Saman in stylized other voice): Ni är så civiliserade!

Saman: (directed to Amina in stylized other voice): Ni djur!

Hayder: Det har varit så; långt innan Ayatollahs (referring to the Islamic revolution).

v

Två chilenare från röda, ah dom va kusiner
Dom bruka importera abiat som det apelsiner
Alla va med, hela familjen fyllde kassaskrinen
Nu dom torska ja, så mormor åkte Aina bilen

Två juggar bor i villa nu, Soran å Goran
Dom kunde bänka 150 utan anabola
Dom har vart med sen åttiotalet, bruka japp Carola
Dom gjorde para innan alla andra hade några

Arash från Hallonbergen var en svenne-iranier
Han kunde snacka sönder suedi sedan mellanstadiet
Bästa säljaren, så firman gjorde fem miljarder
Nu han bor i Strandvägen, hajar i akvariet

(STOR 2016:1:29)

Chapter 3

SOCIAL DIALECTS IN STOCKHOLM

Linguists and non-linguists would generally agree that four social dialects circulate in Stockholm today: (1) the standard unmarked *rikssvenska* (Realm Swedish), (2) traditional Stockholmian *ekensnack* or *södersnack*, (3) upper-class *östermalmsvenska* (Östermalm Swedish), and (4) multiethnolect. Of course, the boundaries between these dialects are fuzzy. Traditional Stockholmian *ekensnack* is considered to be working-class, but King Carl XVI Gustaf uses some of its features, including the stereotyped *Stockholms e* (see p. 47). Older speakers of multiethnolect also use *ekensnack* features. And while middle-class Stockholmians might mostly use *rikssvenska* features, less stigmatized *ekensnack* indicators will occur like pronouncing the vowel in DÖRR like LUDD (Wenner 2010).

Here, however, I treat these dialects as bounded entities for the sake of providing the reader with a foundation in Stockholm dialectology. Table 3.1 outlines the four social dialects, starting with an explanatory label, followed by known academic terms, followed by popular metapragmatic descriptions. The following subsections describe 'ideal' renditions of each.

1 Standard Swedish

As is the case elsewhere, the unmarked standard variety in Stockholm is the home variety of the middle classes. Swedish began to standardize in the 16th century following a grad-

Dialectological Archetypes	Academic terms	Lay terms (neutral & pejorative)
1. Standard variety	rikssvenska (Realm Swedish), standardsvenska (Standard Swedish), centralsvenska (Central Swedish)	‘normalt’ (normal), ‘utbildat’ (educated), ‘utan dialect’ (without dialect)
2. Traditional working-class variety	same as lay terms	stockholmska (Stockholmian), ekensnack (‘Eken’ talk), södersnack (Southern talk), söderslang (Southern slang)
3. Upper-class variety	överklasssvenska (upper-class Swedish)	östermalmssvenska (Östermalm Swedish), lidingössvenska (Lidingö Swedish), ‘saltis’ (from Saltsjöbaden), ‘fint’ (posh), ‘snobbigt’ (PEJ: snobby), ‘bögigt’ (PEJ: gay)
4. Multiethnolect	multietnolekt (multiethnolect), svenska på mångspråkig grund (SMG, Swedish on multilingual ground), förortssvenska (Suburban Swedish)	rinkebysvenska (Rinkeby Swedish), förortssvenska (Suburban Swedish), miljonsvenska (Million Swedish), ‘brytning’ (accent), blattesvenska (PEJ: Paki Swedish), ‘blattigt’ (PEJ: Paki-ish)

Table 3.1: Four social dialects in Stockholm along with their various names.

ual emergence of printed books. *Rikssvenska* (Realm Swedish) took much of its form from emerging orthography, and the orthography was shaped by the variety spoken around the capital city, Stockholm.

> The standardization of the written language [...] [was] marked by the appearance of the Swedish translation of the New Testament in 1526, in a period when the state was consolidated under Gustav Vasa, and when Lutheranism was introduced as the state religion. The codification of the chancery style was strongly influenced by the Svea dialects, in part as a conscious effort to get away from previous Danish influences. The first complete translation of the Bible (*Gustav Vasa's bibel*) was published in 1541, based on the language of the Stockholm district with a fair degree of orthographic standardization. (Riad 2014:1)

Today, linguists and phonologists often refer to *rikssvenska* as *Central Swedish*, which refers to the central part of Sweden where Stockholm and Lake Mälare lie. Nyköping, a small town 50 km southwest of Stockholm, is generally said to be where Swedish is the ‘purest’ (Hellberg 2012) and is where Swedish laboratory phonologists usually collect their data (e.g., Eklund and Traunmüller 1997). Figure 3.1 shows a map of Sweden and the Mälare Valley. Nyköping can be spotted at the bottom edge of the map, and Stockholm is in the center.

Like in many other parts of the Western world, the Industrial Revolution resulted in large-scale migration from the countryside to the cities, and this intensified the leveling of regional dialects in Sweden. In the 1970s, regional dialects began to climb in status, coinciding with a push by the State to include them in (the mostly State-run) mainstream media[1] (Thelander 2009:192).

[1]Historically, there was a ‘prestige divide’ between Southern Sweden on one side and Central and Northern Sweden on the other side. This divide has weakened substantially in the past decade or so. Southern Sweden traditionally viewed ‘lundensiskan’ as its standard prestige variety, centered around the university town of Lund. This

Figure 3.1: Sweden and the Mälare Valley (Ezilon 2009).

Despite that improved status of regional dialects, Leinonen (2010) found strong evidence of continued leveling. She analyzed word-list recordings of older and younger speakers at 98 different locations across Sweden and measured the vowel-quality distances between each location. The differences among young speakers across Sweden was significantly smaller than those among older speakers (2010:157–159).

The following subsections include descriptions of vowels and pitch accent. Consonants are not discussed in this dissertation. Subsequent chapters may warrant returning to this section for reference. The coding of internal predictors in Chapter 6 and the conditional inference tree in Chapter 9 both require an understanding of phonological quantity and pitch accent. Chapter 7 requires an understanding of phonological quantity and quality differences among vowels. It also requires an understanding of the quality differences between Standard Swedish long vowels and traditional Stockholmian long vowels.

style incorporates traditional southern features (e.g., the uvular trill /ʀ/) along with some supra-regional upper-class features. For Central and Northern Sweden, Mälare Swedish has been the main prestige form.

		Long Allophone			Short Allophone		
ortho-graphy	phoneme	short-hand	actual	lexical set	short-hand	actual	lexical set
i	/i/	[iː]	[iː]	DIS	[ɪ]	[ɪ]	DISK
y	/y/	[yː]	[yː]	TYP	[ʏ]	[ʏ]	FLYTTA
e	/e/	[eː]	[ɪ͡e]	LETA	[ɛ]	[ɛ̝]	LETT
ä	/ɛ/	[ɛː]	[ɛː] - [æː]	NÄT - LÄR	[ɛ]	[ɛ̝] - [æ]	LETT - SÄRK
ö	/ø/	[øː]	[øː] - [œː]	SÖT - DÖR	[ø]	[ø]	DÖRR
u	/ʉ/	[ʉː]	[ʉ̟ː]	LUS	[ɵ]	[ɵ]	LUDD
o	/u/	[uː]	[uː]	SOT	[ʊ]	[ʊ]	ROTT
å	/o/	[oː]	[oː]	LÅS	[ɔ]	[ɔ]	LOTT
a	/ɑ/	[ɑː]	[ɑː]	LAT	[a]	[a]	LASS

Table 3.2: Central Swedish vowels divided into long and short allophones. They are listed in the following order: orthography, phoneme, the typical shorthand method of writing the allophone, the typical actual pronunciation, and the lexical set.

1.1 Vowels

Central Swedish has 17 vowels. The general consensus is that nine vowel phonemes occur in short and long allophones; two of the short allophones have merged, rendering 17 in total. This is shown in Table 3.2. According to Kuronen and Leinonen (2001:137), short vowels are on average 67% the duration of long vowels, varying between 58 and 78%. This was assessed on a spoken corpus (sample size unclear). Helgason, Ringen, and Suomi (2013:537) found that Central Swedish short vowels vary between 54 and 64% the duration of long vowels, assessed on the reading aloud of isolated words.

The short-long typology, however, implies that quantity is the only distinguishing factor. This is not the case. Quality plays a major role in their distinction, so a helpful heuristic tactic is to conceptualize the short-long distinction as something like the *lax-tense* distinction in English or German. Table 3.2 is taken primarily from Riad's (2014) account, but I have made a few adjustments. The allophone of /e/ is represented as [ɪ͡e] instead of [eː] to represent the on-glide that occurs in Central Swedish (Eklund & Traunmüller 1997). I have also included what resembles the English *lexical set* (e.g., GOAT, TRAP, etc.) for Swedish. Lexical sets are not widely used in Nordic dialectology, but they really ought to be. Leinonen (2010) has taken a similar position in her dissertation on vocalic variation in Swedish regional dialects. Instead of bracketed phonemes, she used lexical sets from the SweDia project[2]. It made her work easier to read, so I will use her style here.

Note that in this dissertation I will often use the lexical item as shorthand. For example, I will write 'SÖT is lower' instead of 'the vowel in SÖT is lower'.

According to Eklund and Traunmüller (1997), LETA, SÖT, LUS, SOT, and LÅS are highly diphthongal in Central Swedish, DIS and TYP somewhat diphthongal, and LAT and NÄT

[2]SweDia: Phonetics and phonology of the Swedish dialects around the year 2000, (Engstrand, Bannert, Bruce, Elert, & Eriksson 1998-2003)

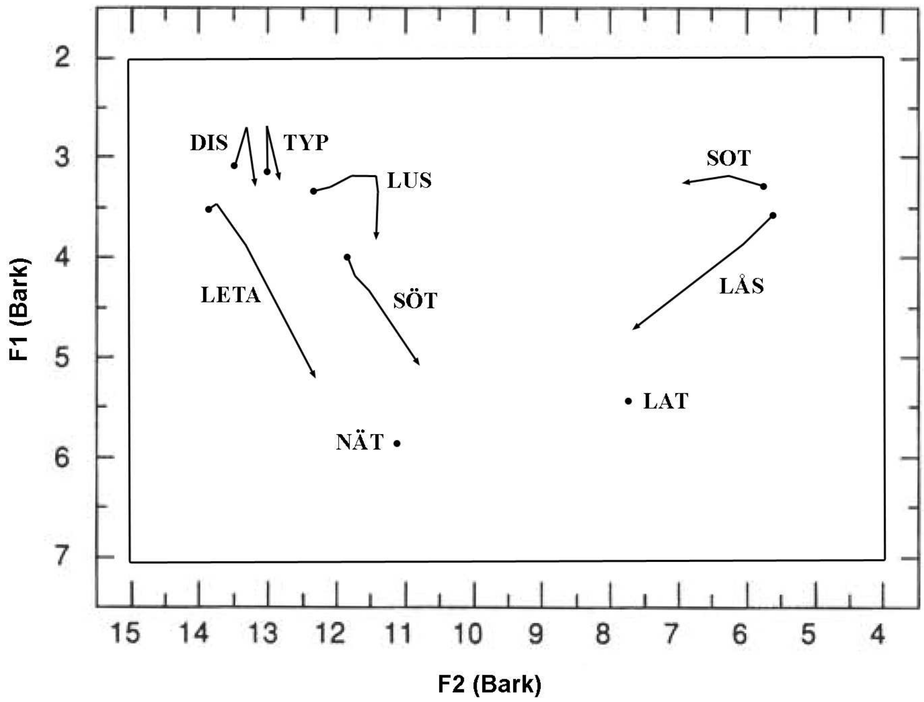


Figure 3.2: Phonologically long vowels in Central Swedish (reproduced and adapted from Eklund and Traunmüller (1997:11)).

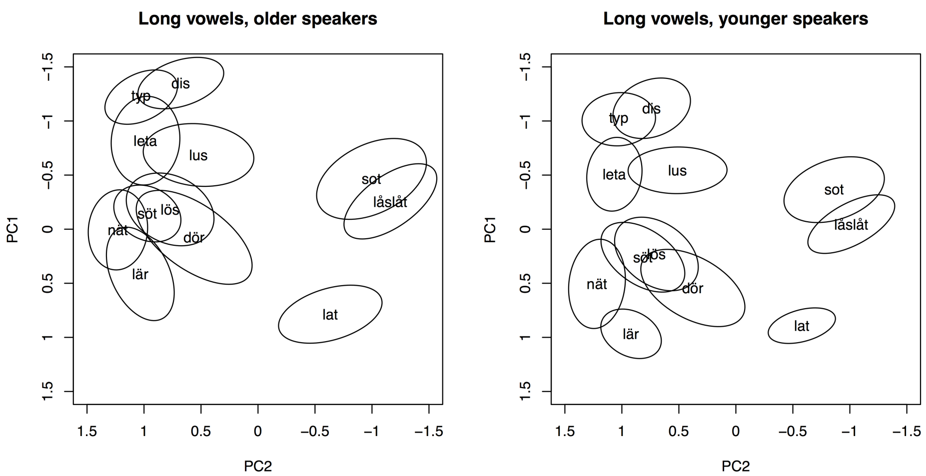


Figure 3.3: Phonologically long vowels of older and younger speakers in the PC2/PC1 plane. The one-standard deviation ellipses are drawn based on the average PC values of the two speaker groups at each site measured at the temporal midpoint of each vowel. (reproduced from Leinonen 2010:165).

	SMALL accent (*usually ω stress*)	BIG accent (*usually ϕ head*)	Example
Accent 1	H L*	L* H	1ˈcir.kus (circus)
Accent 2	H* L	H*L H	2ˈkonst.när2 (artist)
Accent 2 *compounds*	H* L	H*L *H	2ˈcir.kus 2ˌkonst.när2 (circus artist)
			2ˈcir.kusˌkonst.närs2ˌnämnd (circus artist committee)

Table 3.3: Description of pitch accents for Stockholm Swedish as described by Myrberg and Riad (2015:116): SMALL and BIG accents 1, 2, and compound accent 2. Examples are shown with orthodox contours. *ω* stands for word accent, and *ϕ* stands for phrase accent.

monophthongal (1997:11). Figure 3.2 contains a reproduction of their long-vowel chart (1997:11). The diphthongal nature of the long vowels is relevant to Chapter 7 when the question of monophthongization in multiethnolect is investigated.

Nordberg (1975) and Leinonen (2010) report that a downward chain shift has been occurring in Central Swedish long vowels for some time. Nordberg (1975) found that speakers in the industrial town of Eskilstuna produced the SÖT vowel like the more open vowel in DÖR. This 'hypercorrection', as he framed it, likely pulled LUS in such a manner that it resembles the old SÖT today. He also found that young speakers were producing NÄT like open LÄR and that this was stratified between young and old speakers. Leinonen (2010) found this same stratification in her study between old and young speakers for both *Standard Swedish* and *speakers in Stockholm* for the NÄT-to-LÄR and SÖT-to-DÖR admergers (2010:162–163). What is more, she found a consistent lowering for ALL front long vowels among young speakers throughout Sweden. This is shown in Figure 3.3, a reproduction from her dissertation. It shows a clear generational difference for long front vowels (2010:165)[3].

1.2 Pitch accent

Swedish is a pitch-accent language with two lexical accents: *accent 1* and *accent 2*. Typologically, *pitch accent languages* are similar to *tone languages*. Every lexeme belongs to an internally-fixed accent category, and there are minimal pairs distinguished by accent alone. This is what separates Swedish stress accenting from, say, English. According to Hayes (1995),

> Pitch accent languages must satisfy the criterion of having invariant tonal contours on accented syllables, since tone is a lexical property. This is not so for pure

[3]Instead of plotting formants, Leinonen uses the first and second principal components (PC1 and PC2, respectively) of a principal components analysis following band-pass filtering for each vowel (2010:26–29).

> stress languages, where the tonal contours of stressed syllables can vary freely, being determined postlexically by the intonational system (Hayes 1995:49–50).

The main difference between tone languages and pitch accent languages is that the tonal differences in a pitch accent system do not usually manifest themselves in unstressed form. Despite the fact that it is usually true for Swedish, Hyman (2006) might call my claim an oversimplification. According to him,

> [...] Pitch accent is not a coherent notion, rather a 'pick and choose' among the properties that characterise prototypical tone vs. stress-accent systems. [...] Because there is no single prototype of pitch accent, it is not possible to give an explicit definition of pitch accent, as one can for tone and stress accent (Hyman 2006:236).

In other words, Swedish is a 'hybrid' that sits somewhere between Chinese and English on the *tone*↔*stress accent* continuum. In Swedish, accent 1-2 minimal pairs are typically indistinguishable when they occur unstressed in fluent speech. When, however, they are prominent, the appropriate pitch must be used or the meaning changes. For example, ˈ^{2}*tom.ten*$_{2}$ means 'the elf', and ˈ^{1}*tom.ten*$_{0}$ means 'the plot of land'. When the words are stressed, their accents provide a surface-form distinction. When the words appear unstressed, they will typically both become ˈ^{0}*tom.ten*$_{0}$, and one has to rely on context to understand which *tomten* the speaker means.

Pitch accent 1 is the unmarked form for monosyllabic words. Pitch accent 2 is the unmarked form for disyllabic words and compound words – including those that link together accent-1 words. Pitch accents occur as either SMALL or BIG.

> In much previous research, these two intonational categories have been referred to as SENTENCE ACCENT or FOCAL ACCENT and WORD ACCENT or just ACCENT [...]. These terms are problematic, as they imply a strong correlation between the form of the accents and their function. Since the introduction of these terms, we have learned that the functions of the accents are quite varied. Many big accents appear on material that is not focused, and not all words have their own small accent. This leads to confusing formulations (e.g. 'focal accents that appear on given material', 'word accents that cue phrasal structures'). The intention behind the terms BIG and SMALL accent is to avoid this form/function problem. These terms should be sufficiently flexible and theory neutral to be useful for the broad research community. (Myrberg & Riad 2015:117)

Table 3.3 illustrates the acoustic properties for the BIG accent 1 and BIG accent 2. The main difference between the two is that the contour of accent 2 is delayed and carries over to the posttonic syllable. As Myrberg and Riad (2015) have argued, there is no perfect form-function relationship between SMALL and BIG. However, SMALL accents are commonly found on word accents, and BIG accents on phrase accents. It is also important to note that all of this research is based on elicited read-aloud speech. As far as I am aware, there are no comprehensive descriptions for Swedish pitch accent that are derived from spontaneous speech data.

2 Traditional working-class Stockholm Swedish

During Stockholm's rapid industrialization at the turn of the 20th century, a new working-class variety emerged whose features became emblematic for Stockholm as a whole. Industrialization triggered large-scale migration from the countryside with most of the new migrants arriving from Svealand and Götaland (Kotsinas 2001). These new industrial workers settled in the then peripheral, now central, neighborhoods of Vasastan, Kungsholmen, and Södermalm (Kotsinas 2001). After several generations, the rural features that they brought with them koinéized into a new variety (Kotsinas 1988b:144–145). As these new generations were born, the dialectal variants that survived came to index 'streetwise worker' instead of 'country bumpkin' (Young 2018c:58). The symbolic association of the variety with Stockholm's industrial south side, masculinity, and the football club *Hammarby IF* echos the dynamic of London's Cockney. It is therefore surprising that sparse phonetic research has been conducted on it.

Ekensnacket, aka *Södersnacket*, was spoken in Stockholm for most of the twentieth century until other speech styles began to dominate (including multiethnolect). According to Öqvist (2010), the two names for the variety are interchangeable. The first translates literally into 'Eken talk', referring to Stockholm's nickname as *Eken*. The later term translates literally into 'Söder talk', referring to the neighborhood Södermalm where most of Stockholm's working class lived during the middle twentieth century (see pp. 35–37).

Bergman (1946) has speculated that the phonology of the variety most closely resembles those of the old Södertörn and Uppland dialects. However, Öqvist (2010) reminds us that most of what is known about Södersnack phonology is impressionistic and based on stereotypes. Bergman (1946) and Ståhle (1975) both offered impressionistic accounts for three vowel stereotypes ascribed to older working-class Stockholmers. The vowel in NÄT is merged with LETA ([ɪeː] for /ɛː/) such that both become [ɪeː]. This is known as the '*Stockholms e*'. The vowel in LAT becomes like LÅS ([oː] for /ɑː/), and, in turn, the vowel in LÅS is shifted upward to become like SOT ([uː] for /oː/), resulting in a full merge between LÅS and SOT for the most vernacular of speakers. These are catalogued in Table 3.4, which is an adaptation of the table in Öqvist (2010:254). Most modern literature on traditional working-class Stockholmian has republished Bergman (1946) and Ståhle's (1975) findings, offering no new analyses (Kotsinas 1988b; Öqvist 2010).

Kotsinas (1994a), however, is the exception. She is responsible for the only modern sociophonetic study of Södersnack, in which she examined consonants and long vowels of Swedish

Phonemic description	Phonetic description	Lexical set
/ɛ/ merges with /e/	[ɛː] merges with [ɪe]	NÄT merges with LETA
/ɑ/ shifts to /o/	[ɑː] shifts to [oː]	LAT shifts to LÅS
/o/ merges with /u/	[oː] merges with [uː]	LÅS merges with SOT

Table 3.4: *Ekensnack* vowel trends (adapted from Öqvist 2010:254). Left of the arrow represents 'standard' pronunciation, and right of the arrow represents 'extreme' *Ekensnack* pronunciations, implying that variants occur anywhere in-between.

youth from 1989 to 1990. Working-class boys produced 43.7% of the tokens in the NÄT lexical set with the highly diphthongal '*Stockholms e*' LETA vowel ([ɪeː] for /ɛː/). Working-class girls from the southern neighborhoods produced it only 17.2% out of all possible occurrences, compared to 4% for upper-class boys and 1% for upper-class girls (1994a:331).

In addition to these long-vowel transformations, the vowel in DÖRR is often pronounced like LUDD ([ɵ] for /ø/) in Stockholm (Bergman 1946; Ståhle 1975). There is some evidence that this feature is spreading to other parts of Sweden (Wenner 2010).

No research has been conducted on pitch accent and prosody in Södersnack. However, the '*Stockholms e*' variant [ɪeː] for NÄT and [oː] for SOT are much more diphthongal than their Central Swedish counterparts (Eklund & Traunmüller 1997). This, in turn, could result in rhythmic consequences, such as higher intervocalic alternation, which is an important point of investigation in Chapter 7.

3 Upper-class Swedish

In some nation states, the standard variety may be interchangeable with that spoken by the upper classes. This is not the case in Sweden where upper-class speech has always been a distinct marker of exclusivity. As far as I am aware, the only other scholar researching upper-class Swedish is Öqvist (2015). Just as Sweden's upper class is elusive, even less is really known about their speech. Bergman (1946) provides an early and colorful description of stereotypical upper-class pronunciation.

> Who has not heard gentlemen of high-pedigree Stockholm families speak of the city councilman who lived on *Vasagatan* with an open and light *a*, which seems to oppose itself against the definitively closed and common *a*-sound? Even the *sj*-sound seems to be socially graded. The more fronted *sj*-sound that we have in *forsen* and *härs* and *tvärs* has in noble speech [...] more and more pushed out the old *sj*-sound in *sju* and *sjåare*[i] (my translation, Bergman 1946:102).

Here, Bergman is referring to the somewhat delightful iconicity of two linguistic stereotypes in Stockholm, the vowel in LAT, and the palato-alveolar fricative in *sju*. For the former, the fronted [aː] is associated with the upper class, the intermediate [ɑː] with the middle class, and the backed [ɒː] or [oː] with the lower class. For the latter, the fronted [ʂ] is associated with the upper class, the intermediate [ɧ] with the middle class, and the backed [x] with the lower class. Satisfyingly, Stockholm's multiethnolect adds to this iconicity cline by producing [χ] as its variant.

Dual LAT variants in Stockholm – front for the upper class and back for the lower class – are also cited in Öqvist's (2010; 2015) observations. However, in Kotsinas' (1994a) study, the data indicated that the closed back variant for LAT was disappearing and that the open front variant had been re-appropriated by girls from *both* lower and higher social classes. Girls from the southern working-class suburbs, and girls from the wealthy northern suburbs in Stockholm used fronted LAT between 57% and 77% of the time. In contrast, boys from both social classes produced the intermediate variant 80% of the time with no class distinction. Only 5% of the boys' variants were backed while 15% of the variants were fronted; here, there was no class distinction either (1994a:333). This implies a not-unusual connection between upper-class indexicality and feminine indexicality (Eckert 1989b; Steadman 1938; Veblen 1931).

Kotsinas (1994a) also found that upper-class adolescents were using an innovative variant that she had not heard before, namely the open (lowered) LÄR vowel for NÄT. 35% of the boys' and 30% of the girls' NÄT variables were pronounced with the open LÄR variant (1994a:331). Gross, Boyd, Leinonen, and Walker (2016) later found that this feature had matured in Stockholm and was widespread in the speech of teenagers there.

We can now turn to one of the most pervasive upper-class stereotypes circulating about Swedish today: the affricated and damped variant for DIS and TYP. It is often referred to as the '*Viby i*' or '*Lidingö i*'. Lidingö is an upper-class island suburb to the northeast of Stockholm, and Viby is the rural village where the variant supposedly first emerged. Bergman (1946) does not mention the Lidingö *i*' or the Viby *i* in his early accounts of upper-class Stockholm speech. Its absence implies that it might be a newer feature in upper-class speech. This is the position taken by Öqvist (2015) who notes that the only member of the Royal Family who uses the feature is its youngest member, Princess Madeleine.

It is a typologically rare sound found in only a few other languages– e.g., Bora (Parker, Mielke, & Pennington 2018) and Turkish (Björsten & Engstrand 1999). In Gothenburg and much of the Swedish west coast, the variant is more widespread and lacks social charge. Karlgren (1915:295) used [ʅ] to denote the original Viby variant. Riad (2014:21) uses [iː^{z}] and [yː^{z}] for DIS and TYP, respectively. Björsten and Engstrand (1999) and Gross and Forsberg (2017) represent both DIS and TYP with [ɨ]. This is because Björsten and Engstrand (1999) found evidence that the variants were merged in Viby, and Gross and Forsberg (2017) found evidence that the variants were merged in Gothenburg among younger speakers.

In Stockholm, the only sociolinguistic data on this variant is from Kotsinas' (1994a) study. She found that neither working-class boys nor working-class girls used the damped DIS or TYP at all. In contrast, upper-class girls used it about 60% of the time for DIS and 80% of the time for TYP. Boys used it circa 40% of the time for both DIS and TYP (1994a:329). Of all the stereotypically upper-class variants in Kotsinas' study, this one seemed to be the most widely used by upper-class youth.

Riad (2014) has claimed that the variant is spreading and losing its social charge. This is supported by Gross and Forsberg's (2017) findings in Gothenburg. Öqvist (2015) argues that its absence within the royal family points to the fact that it is actually a feature of the *upper-middle class*. Popova (2007) inadvertently provides support this in her popular ethnography of Stockholm's upper class. One of the aristocrats in her book presents the variant as somewhat nouveau riche:

> Popova: How do you tell whether someone is upper-class?
>
> David: [...] The most trivial [features] are the *i*'s. If they are exaggerated, they become a joke[ii]. (my translation, Popova 2007:79–80).

Returning to the review I outlined on Stockholm's new upper class (pp. 31–35), a more clear explanation is that the 'true' upper class is so statistically irrelevant that their speech is never heard and therefore seldom targeted. On the other hand, the ostentatious new elite, while perhaps technically upper middle-class according to canonical understandings of class, is actually the highest-status group that is visible. And this is the group popularly associated with the Viby *i*.

No research to date has been conducted on pitch accent, rhythm or other prosodic features in upper-class Swedish.

4 Swedish multiethnolect

In this section, A substantial amount of space will be dedicated to reviewing the naming, history, development, and features of Stockholm's multiethnolect

4.1 On naming

There has been some debate on what terms best designate the innovative speech practices of Europe's emerging racialized working class. In the case of Sweden, Bodén (2007), Ekberg (2007) and Svensson (2007) at first advocated for *Swedish on Multilingual Ground* but have since distanced themselves from the term. In my earlier work, I promoted the lay term *Suburban Swedish* (*förortssvenska*, Young 2014), which Bijvoet and Fraurud (2016) also later adopted[4]. In my later work, I appropriated the lay term *ortenspråk* (hood language, Young 2018c) in the same way Wiese (2012) has done with *Kiezdeutsch*.

Cheshire, Nortier, and Adger (2015) and Quist (2005, 2008) promote Clyne's (2000) *multiethnolect* while Eckert (2008) has been critical, cautioning that we risk reifying otherwise multifaceted speech practices. She also argues that the term be expanded to refer to the speech

[4] They, however, use this term more narrowly than I did. Whereas my definition aimed at including all possible constellations of suburban features, their definition includes only the speech that has shed all slang forms and includes only phonetic variants (Bijvoet & Fraurud 2016:20).

of dominant groups like white Californians (2008:29) in cases where they themselves consider ethnic distinction just as important as the subordinated groups do. Cheshire et al. (2015) also suggest that *multiethnolect* be used "with caution, recognizing that ethnicity may no longer be a relevant social factor and without wishing to reify a way of speaking that is dynamic and far from focused" (2015:5). This has also been Fraurud's concern (personal conversation, 2014). Benor (2010) offers *ethnolinguistic repertoire* as an alternative. *Repertoire* shifts the focus to specific phonetic variants that index X ethnicity, which is an improvement over *lect*, a term that insinuates that those variants all constellate in the same way.

Much of the problematization of *multiethnolect* arises from a poststructural tradition that seeks to dig deeper and understand how the individual components – usually phonetic variants – are symbolically associated with specific social constructions like ethnicity. The concern with using *multiethnolect* is that its dynamic constitution will be forgotten. Another concern is that the gradient distribution of features will be overlooked. These are certainly valid concerns. An excessive reification of the phenomenon can draw focus away from the fact that the innovations associated with multiethnolect are part and parcel of the systematic variation in the surrounding speech community.

The danger of reifying dynamic constructions lies in many other places as well. Importantly, this includes the very construction of ethnicity itself. Wimmer (2008) argues that ethnicity is momentarily constructed and reconstructed[5] much like lects are. Just as the constellation of variants vary between speakers of supposedly the same multiethnolect, so do the constellation of body artifacts between individuals of supposedly the same ethnicity.

According to his account, the lines between ethnicities can be blurred in one moment and crystalized in another. These momentary strategies of closure and differentiation happen at the individual level, and language use is a core example of one such strategy (Wimmer 2008). He cites Bailey's (2000) ethnography in which a Dominican-American adolescent "situationally emphasizes his black, Hispanic, or American identities" at different times (Wimmer 2008:999). The student's heritage ethnicity does not fit into the USA-mandated triad 'White', 'Hispanic', 'Black', so he uses one set of linguistic forms to mark his belonging to 'Hispanic' in one encounter and another set to mark his belonging to 'Black'.

In this respect, the preference for ethnolinguistic *repertoire* over ethno*lects* may certainly address the reification of speech, but it does nothing for the reification of *ethno*. It draws attention away from the fact that ethnicity is constructed through stylistic practice, one component of which is performative in Austin's (1962) definition of the word. Wimmer (2008) also notes that ethnicity as a construction is constrained by how 'thick' (established) the ethnicity is in that particular community (2008:1003). Analogous to that is the establishment of dialects as coherent constructions. Just as features begin to focus in a community, so too does their variation tighten and approach a more cohesive construction like 'multiethnolect'.

Much how the Dominican-American teen in Bailey (2000) acquiesces to the imposed paradigm of race – and therewith inadvertently contributes to the erasure of his own ethnic

[5] This is a liberal expansion of Weber's (1978 [1922]) constructivist definition of ethnicity.

heritage – this is what is happening in Europe among second-generation migrants. As sundry variants focus into a relatively cohesive constellation, they no longer are actually multiethnic in indexicality. Rather, they are part of an ethnic erasure that is replaced with a racialized dyad that encapsulates allochthonity/ non-whiteness/ Other-ness on one side and autochthonity/ whiteness/ sameness on the other (see, e.g., Cornips 2019, forthcoming; Cornips and de Rooij 2013).

4.2 Early years: *Semilingualism* and *Rinkeby Swedish*

In 1988 Kotsinas identified a youth variety in the Stockholm suburb of Rinkeby that she called *Rinkeby Swedish* (Kotsinas 1988a). Her work, however, was preceded by a large amount of popular (and some academic) material on a 'troublesome' linguistic situation in Sweden's new immigrant suburbs. From the 1970s through the late 1980s, the tabloid Media ran numerous articles on immigrant youth alongside pseudoscientific references to *semilingualism*, a highly controversial term coined by Hansegård (1968). The alarmist articles interwove themes of criminality and social depravity with multiethnic neighborhoods, their youth, and youth language.

One such article in the tabloid *Aftonbladet* ran a concern piece in 1978, shown in Figure 3.4, titled The picture of Sweden 1978: Born to become dishwashers, janitors and token booth clerks.

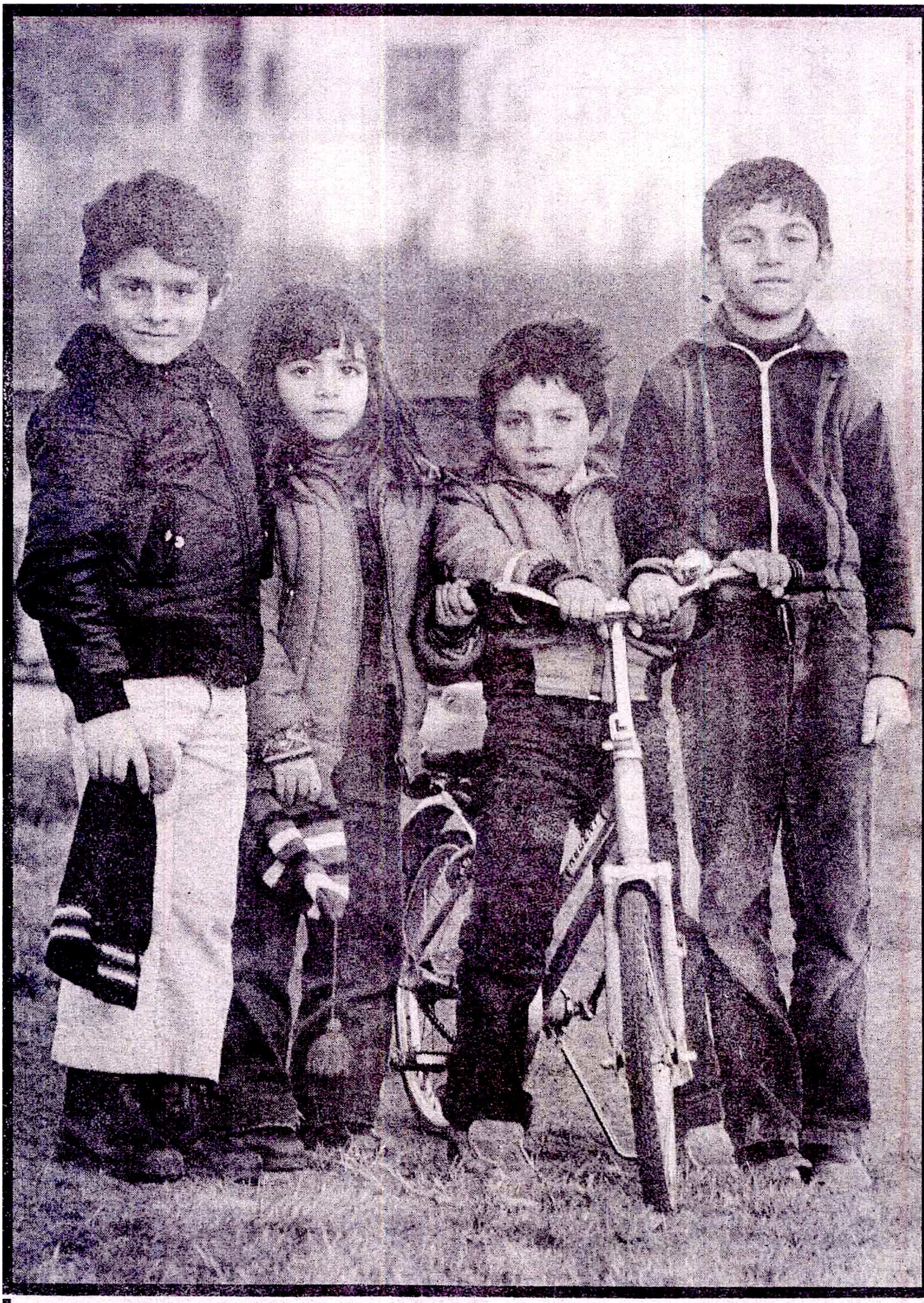
> Mustafa, Yurdagül, Murat and Servet will probably become semilingual. The risk is great that they will not have teachers who know Swedish. But they won't learn Turkish properly either. Two hours of home language instruction is not enough when the Swedish infiltrates in. They will be without a language and without a culture. Falling between the cracks[iii]. (my translation, my emphasis, Lemberg 1978)

Another headline read This is why 16-year-old Petro cracked in Immigrant Sweden (Därför knäcktes 16-årige Petro i invandrar-Sverige). It reported on a young Greek-Swede's trouble with the law.

> Petro is one of these semilingual immigrant youth who stray around in our society today, without any real primary education, without a chance for a job, without a real identity, without a future actually[iv]. (my translation, my emphasis, Svedgård 1980)

The article mentions that his parents immigrated to Sweden and moved to Rinkeby when he was four. Given what we know about the Critical Period of language acquisition (Abrahamsson 2012:209, Hyltenstam 1992:363), such an early age of onset rules out learner interlanguage for even the most extreme outliers.

Dagens Nyheter reported with the headline The children are speaking their own language - that no one else understands (Barnen talar egna språk - som ingen annan



BILDEN AV SVERIGE 1978

FÖDDA TILL DISKARE, STÄDARE OCH SPÄRRVAKTER

Vart fjärde barn som föds i Sverige i dag är invandrarbarn. Men vilken framtid erbjuder vi Mustafa, 7 år, Yurdagül, 6, Murat, 5, och Servet, 7?

Redan nu kan vi säga var de finns om 15 år. Bakom diskmaskinen, vid löpande bandet eller på något annat dåligt betalt eller smutsigt jobb.

Invandrarbarnen i vårt land har ingen framtid. Vi ger dem ingen framtid. I år går 10 000 invandrarbarn i första klass. En fjärdedel av dem börjar på gymnasiets tre- och fyraåriga linjer. En mycket liten del fullföljer sina studier.

Mustafa, Yurdagül, Murat och Servet har turkiska föräldrar. De bor i Västra Flemingsberg i Huddinge utanför Stockholm. Där är nästan 40 procent invandrare.

En invandrarlärare säger:

— Det verkar som om det svenska samhället vill att invandrarna ska fortsätta att vara diskare. Skolan och samhället ger inte barnen en chans.

Mustafa, Yurdagül, Murat och Servet blir troligen halvspråkiga. Risken är stor att de får lärare som inte kan svenska. Men de kommer inte att lära sig turkiska ordentligt heller. Två timmars hemspråkundervisning räcker inte när svenskan tränger sig på.

De blir utan språk och utan kultur. Hamnar mittemellan.

Här står Mustafa, Yurdagül, Murat och Servet. Blivande diskare, städare och spärrvakter.

Bild: ULLA LEMBERG

Figure 3.4: The term *halvspråkig* (semilingual) in the media. The picture of Sweden 1978: Born to become dishwashers, janitors and token-booth clerks. Page from the tabloid *Aftonbladet*, 1978 (Lemberg 1978).

förstår). The article reports on a preschool in Rinkeby where over 30 languages are spoken, 'causing' some children to become apathetic and other children to become aggressive and loud. Many are reported to have speech impediments and mental disorders (Jerkert 1981). A year later, *Aftonbladet* published an editorial piece by the famous Kurdish-Swedish author and human-rights activist Mahmut Baksi (1982) where he condemned the state of schooling in suburbs like Rinkeby and claimed that it was the cause for "envy and feelings of inferiority [among immigrant youth] that were creating contempt and hate" for Sweden (mindervärdeskänslor som skapar förakt och hat).

The avalanche of articles about semilingual youth framed a 'problem' for experts to come in and resolve. This task was put on researchers in Bilingualism, most of whom continued to refute the existence of semilingualism and push for Sweden to keep its progressive bilingual education policy[6]. Amid this discourse, anthropologist Robbins Burling (1981) penned an article in which he criticized the persistent use of the term *halvspråkig* (semilingual). He pointed out a number of similarities between the narrative in Sweden on immigrant youth language and the narrative in the United States on African-American Vernacular English. While acknowledging some key differences between the two phenomena, including Sweden's progressive heritage-language program, he declared the ideologies in both societies to be very similar.

> When Swedish observers claim that the language deficiencies of immigrant children render them incapable of adequate expression in any language, they remind me of Bereiter and Engelmann who believe that many black American children lack an adequate language. Both in Sweden and in the United States, people seem to be capable of inventing theories that rationalize their attitudes about language. (Burling 1981:46)

It was his assessment that policymakers in both places, energized by their misconceptions about youth language, overestimated the power of schools and underestimate the power of families and peer groups on how young people speak (1981:45). He made an appeal for reason and emphasized that there was no scientific evidence that the children of immigrants were linguistically deficient.

It would appear, however, that Burling's plea for reason either did not reach the general public or was unheeded. In my database search of print media, the query *halvspråkig* rendered hundreds of examples, peaking at 26 in 1979 and declining by the early 90s (shown in Figure 3.5. Earlier articles were often in reference to the children of Finnish immigrants in industrial towns throughout Sweden. Later articles, toward the beginning of the 1980s, began applying the term to the children of Greek, Yugoslav, and Turkish immigrants. According to Stroud (2004),

[6] Sweden's initiative for early bilingual education started early and was (and remains) progressive for its time. It guaranteed universal, free, individualized heritage-language instruction for all L2 Swedish children for a set number of hours per week.

Figure 3.5: Search results for the term *halvspråkig* (semilingual) in Swedish print media

> despite the fact that it was subjected to a great deal of critique, especially with respect to the foundational conceptual assumptions on which it rests [...], semilingualism has been felt by many to be immediately appealing. Part of the reason for its tenacity is most certainly because of the perceived 'self-evident nature' of the complex of interrelated ideological suppositions upon which it rests. First of all, as with Rinkeby Swedish, the concept assumes the overall and sovereign importance of the mother tongue – a prerequisite for an individual's cognitive development and emotional harmony (Stroud 2004:207).

Fraurud (2003) has argued that the discourse surrounding semilingualism and Rinkeby Swedish are codified ways by which diversity (mångfalden) is framed as a problem (2003:79). Similarly, Stroud (2004) argues that the discourse surrounding semilingualism and Rinkeby Swedish are part and parcel of the same marginalizing process. "Just as with Rinkeby Swedish, one of the implications of labeling an individual as semilingual is to locate him/her outside of the legitimate market. In this case, this is accomplished by divesting the immigrant speaker of a voice or speaker role" (Stroud 2004:2007).

Evidence from my literature review of the media supports his argument. The same type of alarmist news articles continued – only with *semiligualism* replaced with *Rinkeby Swedish*. Figure 3.6 shows an example of such an article from 1991 that reads HELENA, 7 YEARS – THE ONLY SWEDE IN HER CLASS (Helena, 7 år – klassens enda svensk). The article refers to Kvarnbyskolan in Rinkeby, the same school that one of the participants in this study attended (he would also have been in first grade in 1994, three years after Helena). A teacher remarks that Helena's linguistic target has shifted from her caregivers to her peers at Kvarnbyskolan:

> It was after Helena had been attending for four months that problems began to emerge. "Suddenly I became aware of how she was speaking. She used incorrect word order, she spoke 'Rinkeby Swedish'. In Kindergarden this had never been a problem [...] I contacted the school nurse and got her into a support group for

Helena, 7 år — klassens enda svensk

Av ARNE WIRÉN

Helena, sju år, är unik.

Hon är svensk.

Av 49 förstaklassare i Kvarnbyskolan i Rinkeby i Stockholm är hon den enda vars båda föräldrar är födda i Sverige.

Helena har blivit minoritet i sitt eget land. Nu får hon hemspråksundervisning för att behålla det svenska språket.

För skolans rektor Birgitta Bengtsson är det här problemet inget nytt.

— Sådan är vår situation i Rinkeby, konstaterar hon.

Helenas lärarinna Anna-Lena Rannelid har arbetat i Rinkeby i fyra år. Det här är första gången hon fått en helsvensk elev.

Rinkeby är landets mest invandrartäta stadsdel.

— När jag började här 1972 var 36 procent invandrare. Ett pittoreskt inslag, tyckte vi då, säger Birgitta Bengtsson.

— När det var 50 procent invandrare sa vi att nu klarar vi inte mer.

— Nu är det 93 procent…

När skolan börjar flyr svenskarna

Den nybildade stadsdelsnämnden i Rinkeby har tvingats slå fast som mål för framtiden att "svenska barn ska ha samma möjlighet till inlärning av en bra svenska som barn i mindre invandrartäta områden".

— Svenskar bor här bara tills barnen ska börja skolan. Då flyr de, säger Helenas mamma Ulla-Britt Maxe.

— Man får inte prata om det. Då blir man kallad rasist. Men det är inte det det är frågan om. Jag tycker om människorna här. Det kan bara inte vara rätt att flickan ska få sämre svenska av att gå i en svensk skola.

Det var när Helena gått fyra månader i skolan som problemen dök upp.

— Plötsligt blev jag uppmärksam på hur hon pratade. Hon använde fel ordföljd, hon talade "Rinkebysvenska".

— På dagis var det aldrig några problem. Hälften av barnen var svenskar, hälften kurder. Det blev aldrig den här blandningen av språk.

Fick in henne i en stödgrupp

— Jag kontaktade skolsköterskan och fick in Helena i en stödgrupp för de svenskspråkiga barnen, berättar Ulla-Britt Maxe.

Nu får Helena två gånger i veckan träffa sex andra förstaklassare som antingen har en svensk förälder eller är nordiska invandrare och talar full svenska.

— Just nu läser de "Mio, min Mio", berättar rektorn Birgitta Bengtsson. Det vore en omöjlighet i klassen där Helena går. Fem barn talar nästan ingen svenska alls, sex talar hjälpligt, resten talar men med grammatiska fel och uttalsfel.

— Det anmärkningsvärda är att många av barnen är födda i Sverige och ändå knappt pratar svenska när de börjar skolan.

— Man känner vanmakt. Vi pratar svenska med dem i skolan. Sedan går de hem till sin hemspråksmiljö och pratar det andra språket ända tills de kommer hit nästa dag.

— Följden blir att undervisningen i andra ämnen blir lidande.

Foto: BJÖRN LUNDBERG

Elva modersmål finns representerade bland de 17 barnen i Helenas klass i Rinkeby. Helena är den enda som talar enbart svenska. De andra pratar kinesiska, grekiska, bengali, turkiska, arabiska, spanska, syrianska, wollof (talas i Gambia), persiska och engelska.

Figure 3.6: The term *rinkebysvenska* (Rinkeby Swedish) in the media. Helena, 7 years – the only Swede in her class. Page from the tabloid *Expressen*, 1991 (Wirén 1991).

> Swedish-language children", Ulla-Britt Maxe relays. Helena now meets with six other first-graders twice a week who either have a Swedish parent or are Nordic migrants who speak full Swedish[v]. (my translation, Wirén 1991)

The teacher's reference to *Rinkeby Swedish* was coined several years earlier by Kotsinas (1988a). She conducted the first research on multilingual youth language in Stockholm and accompanied this research with a great deal of media outreach. She penned several newspaper editorials and participated in a documentary on *Sveriges Television* where she introduced the term *Rinkeby Swedish* (1985). It has remained in the Swedish national consciousness ever since.

Kotsinas emphasized that the systematicity of the speech style was evidence that it differed from L2-Swedish interlanguage. She suggested that this variety could have been rooted in a creolization process, whereby marked structures were replaced by less-marked structures in areas such as grammatical gender, tense, and prepositional systems (1988a:264), caused by having learner-Swedish as the target (1988a:269). Later, as she continued to analyze her corpus, Kotsinas distanced herself from the creole assessment and introduced the notion of a new working-class sociolect born out of a koinéization process similar to how Stockholm's industrial-era *Ekensnack* evolved amid rural immigration (Kotsinas 1988b:144–145).

Kotsinas offered impressionistic descriptions of the phonology of Rinkeby Swedish. In her earliest articles, she drew attention to its 'jerky' rhythm and attributed it to a reduction in the difference between long and short syllables (Kotsinas 1988a, 1990). She also found a high amount of individual variation. In her words, individual speakers produce "mistake-like" pronunciations that are not repeated by other speakers. This included deviations from the standard Stockholmian vowels in TYP and SÖT, the full realization of the unstressed vowels in ˈ*va.rit* and ˈ*pla.ttor*, and the hypercorrected pronunciation of the elided *d* in *ibland* (1990:264). She also found fewer intra-lexical elisions, so speakers were more likely to say *har inte* instead of *har'nte* (1990:264), which may have contributed to the jerky rhythmic impression.

Kotsinas also acknowledged that some adolescents could style-shift into standard Swedish (1988a:276). In her later work, she also introduced the possibility that children who spoke standard Swedish in their early years were targeting Rinkeby Swedish in their adolescent years (Kotsinas 1994a:313). She cites Hewitt (1986:150–188) who reported that young Black British speakers who spoke an exclusively standard dialect in early childhood adopted patwa slang in adolescence to signal group affiliation.

Later scholarship has criticized the early work on Rinkeby Swedish for its use of the term creole (Milani and Jonsson 2012:45; although c.f. Wiese 2009:803) and for over-essentializing the variety (Bijvoet & Fraurud 2010, 2013). Nonetheless, Kotsinas was the *only* scholar to produce research on multiethnolect in Sweden from 1988 through 2002. It was not until Fraurud (2003) that additional scholars entered the discourse.

4.3 Later years: Style, variation, and *multiethnolect*

The project *Language and language use among young people in multilingual urban settings* (SUF project; Källström and Lindberg 2011) was established in 1999 to investigate the Swedish used by adolescents in the multilingual areas of Sweden's three major cities: Stockholm, Gothenburg, and Malmö. Data was collected from 222 adolescent participants. The project lasted until 2006 and reopened investigation into language variation in Sweden's urban areas after a 15-year hiatus. Over 50 works were produced by the 15 researchers involved. These works generally fall into one of two categories. The first is descriptive, and the second investigates the constructions and paradigms surrounding speech practice.

Most of the descriptive work examined syntactical and lexical phenomena (Ekberg 2007, 2011; Ganuza 2008, 2011; Svensson 2007, 2011; Tingsell 2011) with the exception of Bodén (2004, 2005, 2007, 2010, 2011) who concentrated on phonetics and phonology. It was in this descriptive work that the decision was made to replace the conflating *Rinkeby Swedish* with the more general *Swedish on multilingual ground* (*svenska på mångspråkig grund*, Ekberg 2007:i–ii).

Ekberg (2007, 2011) identified lexico-grammatical traits unique to the speech of multilingual youth in Malmö such as the polyfunctionalism of *sån* (such) (2007; 2011). A key finding was that *sån* had grammaticalized into both a demonstrative pronoun and a new determiner, replacing the standard indefinite article *en/ett* (a,an). In her investigation of the discourse marker *å sånt* (and such), she found that it was often used as a metonym-like modifier. In a phrase like "duscha å sånt" (take a shower and such), the *å sånt* would refer to all the grooming and readying activities that typically occur with showering (2007:66).

Ganuza (2008, 2011) investigated adherence to the subject-verb inversion constraint among youth in Malmö, Gothenburg, and Stockholm. Known as the 'verb-2 constraint', it typically requires the verb to move before the subject when the first word in the main clause is not the subject: "å sen <u>gick</u> jag" (XVS) instead of "å sen jag <u>gick</u>" (XSV), meaning "and then I <u>went</u>". She found that the XSV word order (as opposed to XVS) – a well-known and somewhat emblematic feature of Swedish multiethnolect – actually was only used in 3.5% of all possible occurrences and in 10% of peer-to-peer occurrences (2011:93). However, usage of this feature increased in situations where suburban identity was overtly performed. It was concluded that XSV was a semiotic tool used to mark group solidarity (2011:99).

Svensson (2007, 2011) investigated discourse particles in the speech of multilingual youth in Malmö. She found that *ba*(*ra*) (just) was used the most frequently and functions as a focus marker (2011:224). Pragmatically, *duvet* often signals the limits of clauses, *ju* stresses truth conditions, and *liksom* and *typ* hedge propositions and function as exemplifiers (2011:225). Indexically, *duvet* (you know), *ju* (well) and *liksom* (like) all signal intimacy and concert between speaker and listener (2011:225).

Tingsell (2011) investigated the use of anaphoric binding in the speech of multilingual youth, focusing on the abandonment of the typical reflexive forms *sig*, *sin*, *sitt*. Although she also investigated the non-possessive *sig*, she found the largest variation to occur in instances of possession where *sin* and *sitt* are otherwise called for. An example of this is using "hon gillar hennes hund" (she likes her dog) instead of "hon gillar sin hund". The former *hennes* is the non-reflexive third-person possessive *her* in standard Swedish, and the latter *sin* is the reflexive possessive *her* in standard Swedish. In 12% of the instances that call for a reflexive possessive pronoun in standard Swedish, the personal possessive pronoun was found to be used instead (2011:242).

Bodén (2004, 2005, 2007, 2010, 2011) conducted her analyses on the pool of 102 participants who hailed from the multiethnic neighborhood of Rosengård in Malmö in the southernmost county of Scania. Within that pool of participants, speech samples were taken and played to listeners from the same pool to identify samples that overwhelmingly were identified as *Rosengård Swedish*. Bodén then identified phonetic variants in those samples that deviated from what she recognized to be the Scanian received standard. The following deviating consonants were reported: [tʃ] for /ɕ/ as in 'checka' (check out), trilled [r] for the Scanian uvular /ʁ/[7] (particularly before stressed vowels), and the emergence of the voiced fricative /z/ as an actual phoneme for new loanwords like 'wazzup' (how are you, 'what's up'), 'guz' (girl, 'chick'), 'doz' (backside, 'booty') (2005:2, 2007:26–28, 2010:71, 2011:43).

For vowel duration, she reported that no significant difference existed between Rosengård Swedish and standard Scanian; i.e., long vowels did not shorten, and short vowels did not lengthen (Bodén 2007:29). LUS, typically a diphthongal long vowel, was found to be backed and lowered but was not perceptually deviant from standard Scanian (Bodén 2007:30–31). And interestingly, standard Central Swedish variants were used for LÅS and SOT in Rosengård Swedish instead of the iconic Scanian diphthongs, [ɛo] for LÅS and [eu] for SOT (Bodén 2007:29). One interpretation that I have drawn from this finding is that Malmö's multiethnolect may have spread from Stockholm rather than being an independent phenomenon.

Inadvertently, Bodén (2007) offers evidence in support of Labov's (1993) *Interface Principle*, which proposes that morphosyntax cannot hold indexical information to the extent that phonetic variants do: "members of the speech community evaluate the surface form of language but not more abstract structural feature" (Labov 1993). She concludes that much of what marks speech as Rosengård Swedish *must* be phonetic due to the fact that listeners identified numerous speech samples as such even when those samples lacked marked lexical or grammatical deviations (2007:26). This was something I also found in my earlier perception work on Stockholm Swedish; listeners nearly unanimously identified certain speech stimuli as multiethnolectal even though those stimuli differed only phonetically from the stimuli identified as 'standard' (Young 2014, 2018c).

Aside from the short literature on rhythm that I review in Chapter 5 (p. 103), no other work has investigated the phonetic properties of Swedish multiethnolect.

[7] NB again that Bodén's work did not take place in Stockholm where the apical trill [r] is considered the prototype for /r/. In Southern Sweden, including Scania where Malmö is located, the uvular [ʁ] is the form used for received pronunciation of /r/.

In contrast to Bodèn's descriptive focus, Bijvoet, Boyd, and Fraurud have explored perceptions of multiethnolectal speech as well as the construction of 'Rinkeby Swedish' itself (Bijvoet & Fraurud 2008, 2010, 2011, 2013). In their 2008 study, they found that some listeners will call a broad range of speech styles 'Rinkeby Swedish' and 'accented' while other listeners (presumably speakers themselves of multiethnolect) will only apply those terms to a narrow range (2008:20–22). The diverse range of listener assessments challenges canonical understandings of 'variety'. They conclude that subsequent studies on language use in Sweden ought to take into consideration the complexity of its boundaries and consider how useful the term 'variety' is for understanding the observed linguistic variation (Bijvoet & Fraurud 2008).

Fraurud and Boyd (2011) and Boyd and Fraurud (2010) also argued that the results of the SUF study mandate a reexamination of *who* is included in large-scale urban studies of language. They point out that Labov (1966a) excluded nearly half of the original participants in his study of the Lower East Side because they were non-native speakers of English. In late-modern Sweden, however, these same distinctions would exclude the crux of *most* urban language variation or change.

> [A] large majority of the 222 participants did not fit neatly into dichotomous categorizations such as native/non-native speakers or first/second language users. Most of these young people are in a broad sense multilingual in terms of both their background and their language proficiency and language use, but few of them would, according to prevailing definitions, qualify as native speakers of Swedish nor of their heritage language, nor would they be considered typical non-native speakers or learners. This observation is in itself an important challenge to paradigms presupposing comparison between learners, shifters or bilinguals and a native control group. (Boyd & Fraurud 2010:699).

When they assessed the participants on the following eight bilingualism criteria, 88 unique profiles emerged: (1) age of arrival, (2) age of onset, (3) location of first exposure to Swedish, (4) language use with parents, (5) language use with siblings, (6) language use with friends, (7) language preference, (8) and self-assessed proficiency in their preference language. Monolingual Swedish speakers did indeed cluster into a single profile; the remaining speakers fell into a scattered distribution of various combinations of the above eight criteria. The highest number of speakers who shared a single profile numbered only ten (Fraurud & Boyd 2011:79). They therefore propose that studies of similar phenomena incorporate all-inclusive samples, take into account *intra*-speaker variation, and that they include all appropriate methodologies – qualitative or quantitative – to explain the variation (2010:701).

Eliaso Magnusson and Stroud (2012) take a similar view and question essentialist views of nativeness, non-nativeness, and nativelike-ness. In an ethnographic study of young Assyrian men who worked in a telephone marketing center, they found that the men navigated between nativelike speech when with Swedish customers and non-nativelike speech when with each other. Following Bourdieu (1991), the authors propose a linguistic marketplace paradigm in which nativelike-ness and non-nativelike-ness are *two registers* that have different values in different markets (2012:341).

The problematization of nativeness is an important consideration for works like this dissertation. When a city has a high ratio of L2-Swedish speakers – over 30% in Stockholm and even higher in cities like London or New York – it becomes necessary to include this population in the representative sample, even if it is a regional or traditionally dialectological study.

A later project, *Sociolinguistic Awareness and Language Attitudes in Multilingual Contexts* (SALAM), was structured with this expanded understanding of speaker profiles in mind. The study concentrated on social meaning and the boundaries between constellations of meaning – identified by means of listener perceptions and self-assessments (Bijvoet & Fraurud 2010, 2012, 2016). In a perception test of 12 stimuli, the authors found that listener assessments depended very much on the listener's own linguistic background. The various labels for multiethnolect were applied onto a wider range of stimuli by 'monolingual' speakers. 'Bilingual' speakers, however, were more likely to apply the label to only the most slang-infused stimuli[8]. The authors conclude from this that the selection of features is a process fraught with uncertainty.

> What these young people do can be visualized as navigating a somewhat hazardous sea with islands and rocks that represent constructions of different ways of speaking—to either target and approach (to different degrees) or to avoid—specified along dimensions of class, ethnicity, gender, authenticity, language proficiency, and so on. (Bijvoet & Fraurud 2012:314)

The above quote implies is that the perception and, crucially, reproduction of variants is a precarious business given the fact that each variant carries an underspecified and evolving basket of social meanings.

Missing from the literature is a connection between feature navigation and surface phonetics. That is to say, no study has honed in on the production of a specific phonetic feature across types of speakers and registers. Take, for example, Bodén's finding that the apical trill [r] was occasionally used in place of the received Scanian [ʁ] among Rosengård youth. Boyd and Fraurud (2010) have argued that this sort of sporadic production is constrained by dimensions of class, ethnicity, gender, authenticity, language proficiency, and so on. Similarly, according to Weinreich, Labov, and Herzog's (1968) framework, its sporadic appearance would be a function of speaker, situation, purpose, awareness, and, inevitably, a limitless range of other predictors. What is needed now are empirical accounts of such functional constraints – like age, race, social class, and style – on specific variables such as (r) in Malmö or, as this dissertation seeks to explore, rhythm in Stockholm.

[8]'Monolingual' is the term used for ethnic Swedes with Swedish as a home language, and 'bilingual' is the term reserved for the descendants of migrants. Both terms, however, ignore the exceptional competencies that most Swedish teenagers have in English, which would actually make the former bilingual and the latter trilingual.

Citations in original language

[i]Vem har inte hört herrar av gamla anrike Stockholmsfamiljer tala om stadsrådet som bodde på *Vasagatan* med ett öppet och ljust *a*, vilket liksom opponerar sig mot det betydligt mera slutna och folkliga *a*-ljudet? Även *sj*-ljuden tycks vara socialt graderade. Det främre *sj*-ljudet som vi har i *forsen* och *härs* och *tvärs* har i högreståndsspråket [...] alltmera trängt ut det gamla *sj*-ljudet i *sju* och *sjåare* (Bergman 1946:102).

[ii]Popova: Hur märker du att någon är överklass?

David: [...] Det mest triviala är *i*:na. Om de överdrivs blir de ett skämt. (Popova 2007:79–80)

[iii]Mustafa, Yurdagül, Murat och Servet blir troligen halvspråkiga. Risken är stor att de får lärare som inte kan svenska. Men de kommer inte at lära sig turkiska ordentligt heller. Två timmars hemspråkundervisning räcker inte när svenskan tränger sig på. De blir utan språk och utan kultur. Hamnar mittemellan (Lemberg 1978).

[iv]Petro är en av dessa halvspråkiga invandrarungdomar som i dag driver omkring i vårt samhälle, utan någon rigtig skolutbildning, utan chanse till ett jobb, utan verklig identitet, utan framtid egentligen (Svedgård 1980).

[v]Det var när Helena gått fyra månader i skolen som problemen dök upp. "Plötsiligt blev jag uppmärksam på hur hon pratade. Hon använde fel ord ordföljd, hon talade 'Rinkebysvenska'. På dagis var det aldrig några problem. [...] Jag kontaktade skolsköterskan och fick in Helena i en stödgrupp för de svenskspråkiga barnen", berättar Ulla-Britt Maxe. Nu får Helena två gånger i veckan träffa sex andra förstaklassare som antingen har en *svensk* förälder eller är nordiska *invandrare* och talar full svenska. (Wirén 1991)

Chapter 4

METHODS

Thirty-six men participated in this project. They hail from a stratified sample of social classes and ethnicities, and in 2017 their ages ranged from 24 to 43. Speakers were recorded in either three or four speech styles, depending on the level of participation that they chose. The styles included casual speech between peers, an interview with me, an unprompted reading passage, and a prompted reading passage. Extensive social data was collected in one or two interviews, also depending on the level of involvement that participants chose. The data was transcribed and phonetically time-aligned. The social data was granularized and set in tandem with extracted properties from the time-aligned corpus. From there, I conducted the four analyses in Chapters 6, 7, 8, and 9.

In this chapter I describe the finer details of recruitment, data collection, interview design, ethics, consent, funding, recording, transcription, and phonetic-time alignment. I close the section by theoretically framing two key social concepts that this dissertation relies upon: *racialization* and *social class*.

1 Data collection

1.1 Participant recruitment and inclusion

I presented myself to potential participants as a doctoral researcher from Queen Mary, University of London, conducting research on the daily life and speech of today's Stockholmers. I recruited by means of my own personal and professional networks (see p. 73 on my positionality). I also engaged in the snowball method by seeking out new participants through an existing participant's network.

Thirty-six men from a stratified sample of social classes and ethnicities participated in this project. Due to time constraints – mostly connected to phonetic segmentation – I decided to exclude women in order to ensure that the sample size was sufficiently large along class and ethnic lines (See Section 2). For the same reason, gay and bisexual men are also excluded. The 36 men therefore all self-identify as 'straight' and constitute a subset of 60 Stockholmers whom I interviewed and observed ethnographically between 2015 and 2018.

All participants were born in Sweden except for Antonio (2 yrs), Kevin (4 yrs), and Sohrab (6 yrs) whose age of arrivals are in parentheses (pseudonyms used; see p. 140).

Selection of the 36 participants was based on the aforementioned criteria and willingness to participate in the *main interview*. That is to say, some participants would participate in a peer-group recording and then drop out before I could conduct a main interview with them. The final number of participants who made it to the main interview numbered 37. I removed one because he was unable to read aloud, resulting in 36.

A full list of the participants with their profession, age, and racialized self-identification can be found on page 140.

1.2 *A priori* orthogonality for three key parameters

Correlation matrix:

	age	heritage	SEI
age	1	0.46	0.41
heritage		1	0.27
SEI			1

I recruited participants with an aim for stratified sampling according to three parameters: (1) of Swedish heritage or migrant heritage, (2) socioeconomic index (SEI) as determined by occupation (Ganzeboom & Treiman 2018), and (3) age. Throughout the fieldwork period, I ran regression correlations to ensure that SEI, age, and Swedish/non-Swedish heritage remained orthogonal from each other ($R < 0.5$).

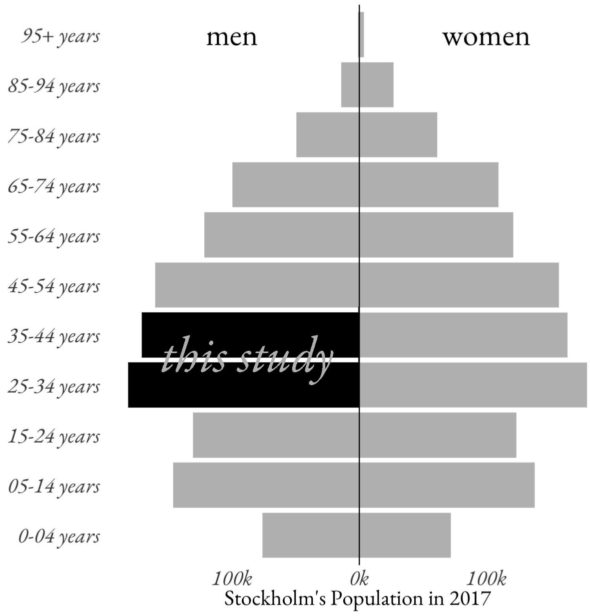


Figure 4.1: Population pyramid of Stockholm according to age and sex for year 2017 (Statistics Sweden). The black bars indicate the cohorts sampled for this study.

1.3 Representativity of the sample

The 36 heterosexual men who participated in this study are only representative of a specific subsection of Stockholm's populace. Women, youth, and the elderly are excluded – as are self-identifying gay and bisexual men. Figure 4.1 contains a population pyramid for Stockholm in 2017 broken down by age and sex. It shows that the age range that I sampled for this study, 24–43, is part of the largest two cohorts within the city's population. It also shows that the focus on male speech excludes a full *half* of this cohort, something I problematize further in the next section.

The group sampled is also considered to be an especially influential demographic in the media. As an example, Lauzen and Dozier (2005) found that 72% of male characters in film were in their 20s, 30s, and 40s. The remaining age groups constituted just 28% of all male roles (2005). This may be a reflection of the fact that all Western societies tend to have a central bulge like in Figure 4.1, which, according to Madsen, Daumerie, and Hardee (2010), stands in contrast to developing nations that have a stark pyramid-like distribution of age cohorts (2010:1). Such youth-dominated pyramid-like structures will have stark implications for how cultural, economic, and linguistic outputs manifest themselves. It is perhaps due to this central demographic bulge in Western cities – and its effect on cultural outputs – that the 20s-to-40s age group is overrepresented in a number of seminal urban studies such as New York (Labov 1966b) and Philadelphia (Labov 2001).

It is therefore important to remember the limits on how much this study can be generalized to other cohorts. While the 20s-to-40s demographic is the largest and perhaps also the most culturally influential, it is nonetheless a numerical minority. Of the 2.3 million residents in Stockholm county, only 352 610 (or 15%) of them are men between the ages of 25 and 45. As it pertains to women between the ages of 25 and 45, they constitute an equally large portion of the population, which makes their exclusion the most problematic. I therefore discuss the implications of this in Section 2.

1.4 Main interview: Adapted sociolinguistic interview

The main interviews were adaptations of Labov's (1972b) *sociolinguistic interview*, particularly the first part. Participants were encouraged to tell stories about their childhood, situations where they had been in danger, superstitious occurrences, and so on, with the aim of eliciting their vernacular style. After approximately 45 minutes to an hour of conversation, I had them read aloud a text by the name of *Cirkusen* (Morris & Zetterman 2011). They then read aloud a word list and minimal pairs (not used in this dissertation). I then asked them to read *Cirkusen* again, but this time like a radio announcer. The details of this portion is discussed in the analysis in Chapter 8 on style-shifting (p. 215)

The interview culminated with a questionnaire on their social background and interests. The goal of this portion was to 'leave no stone unturned' in regards to the participant's social and linguistic background while also respecting time constraints. In this way, my method departs from the classic sociolinguistic interview, which typically weaves in a limited quota of socioeconomic questions. By collecting thorough and detailed data, I aimed to increase the chances of being able to model social class in a more comprehensive manner. I also aimed to increase the chances of being able to qualitatively explain any outliers that the quantitative analysis could not account for. The following outline specifies the questions asked.

1. Classic social factors
   (a) age
   (b) sexuality
   (c) religion
   (d) occupation
   (e) trade union
   (f) income
   (g) housing type and size
   (h) highest-attained education
2. Heritage ethnicity and class
   (a) guardian 1
      i. birthplace
      ii. place where grew up
      iii. age of arrival in Sweden
      iv. ethnicity
      v. first language(s)
      vi. other language(s)
      vii. occupation before immigration (if applicable)
      viii. most recent occupation
      ix. highest-attained education
   (b) guardian 2
      i. birthplace
      ii. place where grew up
      iii. age of arrival in Sweden
      iv. ethnicity
      v. first language(s)
      vi. other language(s)
      vii. occupation before immigration (if applicable)
      viii. most recent occupation
      ix. highest-attained education
3. Assessing bi- and multilingualism
   (a) age of arrival in Sweden, if applicable
   (b) Swedish age of onset
   (c) first language(s) (abbreviated L1)

i. how often does the participant speak L1(s) with:
    A. guardian 1?
    B. guardian 2?
    C. siblings?
    D. friends?
    E. romantic partner?
ii. self-assessment of understanding L1(s)
iii. self-assessment of speaking L1(s)
iv. self-assessment of reading L1(s)
v. self-assessment of writing L1(s)

(d) second language(s) (abbreviated L2)
  i. how often does the participant speak L2(s) with:
    A. guardian 1?
    B. guardian 2?
    C. siblings?
    D. friends?
    E. romantic partner?
  ii. self-assessment of understanding L2(s)
  iii. self-assessment of speaking L2(s)
  iv. self-assessment of reading L2(s)
  v. self-assessment of writing L2(s)

4. Civic engagement and identity politics
  (a) if the participant had to support a political party, which one?
  (b) voted in the last election?
    i. for which party?
  (c) does the participant feel like he is a part of Sweden's society?
5. Speech communities within participant's lifespan
  (a) list of all places lived alongside arrival and departure age
  (b) preschool(s) name, location
  (c) primary school(s) name, location
  (d) secondary school(s) name, location (if applicable)
    i. high-school track (if applicable)
  (e) trade school, university name, location (if applicable)
6. Appearance
  (a) hair style (fade? short? long? manbun?)
  (b) jewelry (pierced ears? chain? rings? gold teeth?)
  (c) tattoos
  (d) cosmetic enhancements (anabolic steroids? botox?)
7. Taste
  (a) favorite musical artists/bands
  (b) favorite athlete(s)
  (c) favorite radio station
  (d) favorite movie(s)
  (e) favorite TV show(s)
  (f) does the participant own a car?
    i. how many?
  (g) favorite car make
  (h) favorite motorcycle make
  (i) favorite shoe brand
  (j) favorite clothing brand
  (k) favorite newspaper
  (l) favorite vacation destination
  (m) what does the participant do in his free time?
  (n) favorite night club or bar
  (o) is the participant part of any social or athletic club?
  (p) does the participant use social media?
    i. Facebook?
    ii. Instagram?
    iii. Snapchat?
    iv. Twitter?
8. Experian Mosaic Market Segmentation survey of interests (Experian 2013)

Not all of this information was used for this dissertation project. Each individual analysis chapter has its own respective Methods section that discusses the social data used and how it is operationalized within the analysis.

1.5 Semi-structured peer group recordings

Some of the participants agreed to partake in a semi-structured peer group recording. I chose this over self-recordings because my earlier ethnographic observations of heterosexual men in Stockholm revealed that they sometimes sit together in silence for long periods of time (even talkative groups will often have a so-called 'Silent Bob' who rarely takes the floor). This appears to be a culmination of two sociocultural facts. First, as Sontag (1969) so humorously describes in her famous essay, *A Letter from Sweden*,

> Talking apparently never ceases to be a problem for the Swedes: a lean across an abyss. [...] Conversation is always in danger of running out of gas, both from the imperative of secretiveness and the allure of silence. Silence is the Swedish national vice (Sontag 1969:26).

Second, traditional notions of Nordic masculinity involve stoicism and measuredness (Ekenstam et al. 1998). These will obviously impede a bountiful collection of speech data, so I decided to take a more involved role. I set up a room, attached lavalier mics and recorders to each speaker, and had them play a game I devised by the name of *Real Talk*. I stayed in a different room while they played the game.

In *Real Talk*, each participant must select a topic from two of three possible themes, illustrated in Figure 4.2. A stack of cards with prompts is placed under each theme:

1. *Everybody's truth*: Cards that contain questions that elicit personal narratives
2. *Myt eller sanning* (Eng. *True or false*): Cards that contain general statements about various hot topics with the intent to elicit indignation and debate
3. *Snacka om snack* (Eng. *Talk about talk*): Cards that contain questions about speech in Sweden and Stockholm with the intent to elicit metapragmatic evaluations

For the sake of space I will not list every prompt, but Figure 4.2 contains a short sample.[1]. The questions under *Everybody's Truth* are inspired by Labov's (1972b) *sociolinguistic interview*. The first question under *Myt eller Sanning* is from Jonsson (2015:114). The remaining are my own. The participants were instructed to go through the stacks of cards and pick the topic they were most interested in. They could address as many topics as they wanted, time permitting. Not all topics were given to all groups because some of the questions were class-sensitive. For example, the upper-class men in Private Equity were given the topic "Those most popular in school have become adulthood's biggest losers", and this question was withheld from lower social classes. Working-class men were given "Is it fake to talk proper Swedish outside the hood?", and this question was withheld from higher social classes.

1.6 Second interview: Social network data

At the end of the main interview, participants were asked whether they would be open to a follow-up interview about their social networks. Twenty-nine of the participants partook in

[1] NB every part of the experiment was conducted in Swedish. The English used on the board game's card base mimics the stylistic use of English in Swedish entertainment and is designed to key that genre for the participants.

EVERYBODY'S TRUTH

Have you ever experienced something supernatural?

Has your life ever been in danger?

Were you generally good or rowdy as a kid? Tell a story.

MYT ELLER SANNING?

Those most popular in school have become adulthood's biggest losers.[a].

People usually don't have any money because they buy stupid things.

The Husby riots[b] were an unreasonable overreaction.

SNACKA OM SNACK

Is it fake to talk proper Swedish outside the hood?

What are *Suburban Swedish*, *Import Swedish*, *Blatte Swedish*, *Rinkeby Swedish*[c]? Are they the same thing?

Can you tell someone is gay just by their speech?

[a] This was taken from Jonsson (2015: 114).

[b] See page 29 and https://en.wikipedia.org/wiki/2013_Stockholm_riots

[c] Sw *förortssvenska*, *importsvenska*, *blattesvenska*, *rinkebysvenska*

Figure 4.2: Game devised for semi-structured peer-group recording sessions

this second interview. The interviews lasted between 30 minutes and three hours depending on the person, and they were conducted orally while I typed the information into a pre-formatted Excel spreadsheet. The participants were asked to list every person they had spoken to in the past month, dividing them up according to job(s), romantic partner, in-laws, family, friends, and hobbies. After the names were provided, demographic information was collected about each contact. Once this was complete, a sociogram was constructed in order to model who within the list knew whom.

I have conducted two literature reviews on social-network analyses that motivate this stage of the data collection. They are located in Chapter 6 (pp. 161–165) and Chapter 9 (pp. 260–264).

2 The omission of women

One glaring gap in this dissertation is the exclusion of women's speech. This is obviously problematic if one wishes to make generalizations about speech in Stockholm because over half the population is missing from the representative sample. Several coinciding limitations led me to the decision to exclude women from the study. I initially set out to study a representative sample of social class and ethnicity across *both* genders. However, as I began to gain insight into the complex social-class system in Stockholm, the required numbers for a representative sample grew.

Midway through my fieldwork, I had interviewed about 14 women and 20 men when I came to realize that I would need approximately 70 individuals in order to have a sufficiently representative sample size. This was untenable for a four-year doctoral project because I also had to allocate time toward constructing a forced aligner. At that point I decided to stop interviewing women and continue with a more expansive social-class sample of men. Men were selected for the following key reason: they already constituted a larger portion of my completed interviews, being easier to access through my own mostly-male social networks.

A second reason is due to the literature that has shown that men generally use vernacular forms more often than women. This trend in the literature, however, is a large aggregated trend within which a number of nuances and exceptions operate. In the ensuing section, I will discuss these nuances in detail.

The nuanced findings regarding gender and the vernacular

According to Labov's (2001), *Principle of linguistic conformity of women*, women produce a lower rate of stigmatized variants and a higher rate of prestige variants than men for variables that are stable within the community (2001:266). This principle is based on findings in Anniston, Alabama (Feagin 1979), Bahia Blanca, Argentina (Fontanella de Weinberg 1974), Belfast (Milroy & Milroy 1978), Buenos Aires (Lavandera 1975), Copenhagen (Gregersen & Pedersen 1991), Detroit (Wolfram 1969), Glasgow (Macaulay 1976), Montreal (Thibault

1983), New England (Fischer 1958), New York City (Labov 1966b), Norwich (Trudgill 1974), Ottawa (Woods 1979), Philadelphia (Cofer 1972; Labov 2001), Santiago, Chile, (Silva-Corvalán 1981), Seoul (Hong 1991), Sydney (Eisikovits 1981), and Tehran (Jahangiri 1980, although cf. Modaressi 1978).

On the other hand, Labov (2001) notes that women are typically the most advanced for incipient changes that are yet to embed themselves in society. This is particularly the case for women from middle social groups (2001:279–293). This tendency toward innovation has been found in Buenos Aires (Wolf & Jiménez 1979), Cairo (Haeri 1996), Detroit (Eckert 2000; Fasold 1969), New York (Labov 1966b), Norwich (Trudgill 1974), Panama City (Cedergren 1973), and Salt Lake City (di Paolo 1988) . This was also found by Kotsinas (1994a) to be the case in Stockholm for the open variant [œː] of the /øː/ phoneme in SÖT and the buzzed variants of the /iː/ phoneme in DIS and the /yː/ phoneme in TYP, respectively.

In his study of sound change in white Philadelphia English, Labov (2001) investigated a full representative sample of social classes and found that "female speakers use nonstandard forms less than male speakers by a factor of 10 to 15 per cent" and that the "coefficients for casual and careful speech are remarkably similar, indicating that it is not merely a special sensitivity of women to the interview situation"(2001:264). This was the trend for most social groups and for most of the variables examined; however, within the lower working class, women and men differentiated the *least*. Importantly also, lower working-class women were more prolific than men for negative concord and the [d] variant of /θ/. Hong (1991) found the same in Seoul, Korea, for the stigmatized tensing of [l] of initial liquids in English loanwords; lower working-class women were as advanced as men. Labov (1966a) found that saleswomen in high-end department stores, when saying the phrase "fourth floor", made a much higher distinction between casual and emphatic speech in the production of [r] than saleswomen in low-end department stores (1966a:49).

Because women seem to be conservative on one hand and innovative on the other, and because lower working-class women seem to counter the trends of their higher-class counterparts, Labov has referred to this as the *Gender Paradox*.

Eckert (1989a), however, argues that it isn't necessarily a paradox at all. According to her, many studies ignore key differences between the gender hierarchies on one hand and class/race/ethnicity hierarchies on the other. Whereas men and women are expected to pair up in lifelong partnerships, the same cannot be said for working-class and middle-class people or white and black people. Therefore, she encourages linguists to take a closer look at the situational meaning of variants within the immediate speech community rather than seeking out global axioms whenever gender complicates social-class effects. For example, the lower-working class "exception" for women may be less due to those women avoiding prestige, per se, and more to do with how romance and femininity are practiced in that subculture. The use of stigmatized variants may be part of the construction of a "street femme fatale" persona that is valued in the romantic and sexual marketplace of lower working-class culture.

More generally, Eckert (1989a) refers to her earlier ethnography of a suburban Detroit high school to remind us that girls are more constrained to symbolic resources than boys are. Whereas boys can maintain the mainstream *Jock* image through athletic prowess and the achievement of feats, girls must do it via the careful curation of "All-American" clothing, affect, and language. Whereas countercultural *Burnout* boys exhibit urban autonomy by traveling to central Detroit alone, girls require male chaperones and therefore rely more on symbolic resources to maintain a corresponding Burnout countercultural persona (1989a:259). When we re-examine the supposed gender paradox of women in light of this, it becomes evident that women simply *do* class more than men. In other words, lower-class women are really "doing" lower class in their prolific production of stigmatized variants. Since lower working-class communities orient the strongest toward stigmatized forms, and since women in these communities are in an even more precarious position than men, women have the strongest incentives to use these forms as a way to harvest symbolic power (Eckert 1989a:256). In a similar fashion, middle-class women are really "doing" middle class by avoiding stigmatized variants and advancing new forms.

Keim (2007) has shown that in the case of ethnolects in Europe, girls and young women are just as in tune with the social constraints imposed on them as the women in the variationist studies discussed earlier. In her longitudinal study of second-generation Turkish adolescents in Mannheim, she found that the girls successfully developed aptitude in a high number of speech registers to adapt to the different and often conflicting norms they were held to. The heavy scrutiny on their behavior – combined with the limitations put upon them as girls – forced them to target symbolic resources as they developed and grew into young women. These resources took the form of mastering numerous registers while preserving their "coarse" (German: grob) way of speaking for insider communication.

–

The reason why I have conducted this review is to spotlight the consequences that the omission of women will have for this study. The prior literature has demonstrated that there are many ways in which rhythm might stratify in Stockholm for the speech of women. Crucially, this literature indicates that it cannot be assumed that the findings for men in this study would be generalizable to women.

Let us now turn to data from this study and examine the gender distribution of the participants' very own social networks. The social-network data implies also that the linguistic picture might be very different for women in Stockholm. The networks of these men are imbalanced from a gender perspective, averaging 22% women and 78% men. Table 4.1 shows a complete breakdown of the percentage of female contacts within each participant's network. Gendered language has typically been discussed in terms of the diffuse networks of women and the more closed networks of men – particularly among the working classes (Bortoni-Ricardo 1985; Milroy 1987, *c.f.* Sharma 2011). What has been less discussed is the isolation of men from women. In many of the participants' networks, the only women mentioned were their mothers and perhaps a romantic partner (some, to my dismay, failed to mention their

Pseud.	racialization	class	age	network women	Pseud.	racialization	class	age	network women
Reman	*invandrare*	0.0	24	19.4	Per	*svensk*	14.7	29	16.7
Malik	*invandrare*	1.1	28	32.7	Rasmus	*svensk*	23.8	34	35.8
Max	*invandrare*	4.4	28	-	Johnny	*svensk*	26.9	30	0.0
Antonio	*invandrare*	7.6	24	-	Sven	*svensk*	34.7	24	-
Hayder	*invandrare*	19.0	25	7.0	Paul	*svensk*	38.2	39	36.4
Dawit	*invandrare*	19.8	26	-	Jesper	*svensk*	38.7	23	0.0
Solomon	*invandrare*	20.1	37	34.5	Richie	*svensk*	52.8	30	40.4
Murad	*invandrare*	21.7	25	35.2	Thomas	*svensk*	60.8	32	38.2
Abel	*invandrare*	25.4	26	-	Nils	*svensk*	64.4	39	0.0
Mateo	*invandrare*	31.6	36	21.7	August	*svensk*	73.5	42	-
Reza	*invandrare*	34.8	25	0.0	Jonte	*svensk*	76.9	43	58.5
Mezdar	*invandrare*	37.5	28	0.0	Martin	*svensk*	82.2	39	21.4
Jocke	*invandrare*	42.2	28	18.6	Gunnar	*svensk*	82.3	37	34.5
Sohrab	*invandrare*	48.6	33	14.6	Joseph	*svensk*	84.1	33	0.0
Shorty	*invandrare*	51.2	26	0.0	Jan-Bertel	*svensk*	86.5	31	37.2
Kevin	*invandrare*	64.0	39	25.7	Johan	*svensk*	88.9	39	0.0
Tarik	*invandrare*	73.4	31	29.0	Jan-Axel	*svensk*	100.0	32	39.7
Majeed	*invandrare*	76.2	30	-					
Parviz	*invandrare*	91.1	24	43.2					

Table 4.1: Percentage of female contacts within each participant's social network

mothers!). Among friends and close contacts, men constituted the overwhelming majority.

In Chapter 6, I show that network qualities are strong predictors of how nPVIV stratifies in the dataset, especially network qualities that relate to class and racialization. This fact, combined with the fact that some participants have no or few women in their networks, leads me to question whether the stratification of rhythm among women would look anything like these results. This is one of the reasons why I have attempted to make my methodology transparent. It is my hope that someone will reproduce this study on female speakers and complete the picture.

3 Research ethics and consent

This project was approved by the Queen Mary Ethnics of Research Committee (Panel D) on January 20, 2017, and assigned reference number QMERC2015/72. All parts of Swedish Law 2003:460 on ethical practices within research that engages humans (om etikprövning av forskning som avser människor, Riksdagen, 2003) were abided.

3.1 My positionality and the ethics of this inquiry

By the time I began collecting data for this project, I had been residing in Stockholm for nearly four years. A native of the United States, I immigrated to Sweden in 2012 after abruptly losing my job in New York City. At one of the lowest points in my life, Stockholm opened its arms to me, and I spent several years working odd jobs in teaching and the restaurant, bar, and nightclub business. During that time I also worked toward a Master's degree in Linguistics at Stockholm University. I was also relatively active in union culture and in some of the local boxing gyms in the northwest suburbs. This is to say that my livelihood was and still is very dependent on the personal connections I have in Stockholm. Much of my social and emotional world centers on individuals who are closely tied to the participants in this project. Therefore, my positionality in regards to the community is one of hybrid insider-outsider. I did not grow up there, but I also did not just show up with a recorder and a notepad. I had grown considerable roots before I set out on this project.

My positionality in terms of power depends on the perspective taken and the participant involved. Swedes enjoy some of the highest living standards in the world. On most measurements, their social and economic wellbeing are higher than in my native United States. Further, some participants had ties to former employers of mine and therefore held more power than I from that angle. Others belonged to symbolically lower-status groups than I – my own background is middle class (with some working-class hysteresis from my father's side). However, due to their Nordic passports, these participants had more access to vital services and privileges than I did. Several participants even remarked on this when the topic of my own heritage came up in conversation. On the other hand, some of these same individuals had experienced racism to a degree that I never have and likely never will.

The confusion surrounding my placement in the social hierarchy may have helped participants feel more secure in our interviews. According to my former supervisor, Kari Fraurud, my hybrid insider-outsider status, combined with my nebulous class expression, may have granted me particular license to tell their story – a license that 'regular Swedes' would likely not have gotten.

I tapped into my own social networks as well as the networks of colleagues and friends to find an appropriately representative and stratified sample of participants. I also used my affiliation with various institutions as a means by which I could approach and recruit participants. This included my barber shop, places of work, gym, frequented bars and restaurants, the local supermarket, and so on.

Some friendships grew out of the interview process because, as I outline above, the questions I ask are intimate. Inevitably, an interview sometimes functions as an event for communion. At its outset, two people might decide to begin a friendship. For example, the interview closes with questions about musical taste, favorite places to go in Stockholm, and interests. In some cases, these resulted in opportunities to socialize. This is not something I shied away from.

Five of the participants in this study (pseudonyms Hayder, Malik, Mateo, Murad and Reman) participated in an earlier study that I conducted for my Master's thesis. One I now consider a friend. Another I see often because he works in the same field as I. Three of them I see on a regular basis because they are close friends, and I, in turn, am close friends with the sister of one. This is generally representative of my ties to many of the participants. Where I am not close to the participant, I typically have close ties to a friend of the participant. In that manner, my network placement is a second or third degree node between all 36 of them.

Social class and ethnicity are taboo subjects in Sweden, so I spent considerable time thinking about how to balance transparency with sensitivity. One concern I had was that the class angle could be interpreted by the participant as robbing him of his own individual identity. Few people are comfortable being placed 'in a box', and the mantra of 'don't label me' carries particular resonance for many. I addressed this concern by bringing it up in those exact words. The interviews finished with a short discussion where I confided in the participant that I wanted to examine social class in a modern way that respects individual differences and identity politics while also acknowledging the limits imposed on one's body by society. In other words, what is the balance between their agency and the sometimes violent reality of their 'box'? In this discussion, a lot of insights emerged from the participants themselves.

Sensitivity to those with less power in society was one of my primary concerns, but I also spent a good deal of energy finding ways to avoid alienating my speakers from the higher social ranks. Stockholm is a small and close-knit city, and some of my right-wing and/or wealthy participants were aware of my leftist political leanings, my involvement with my trade union, and my interest in multiethnolect. It was of concern to me that they would hold back in fear that they would be painted in a bad light in my dissertation. My solution to this was to tell them "and I don't want a politically-correct answer here. Tell me what you really think". This seemed to be effective; I did not get the sense that the elites or right-wing participants held back at all.

Some may argue that distance from one's participants is necessary to maintain objectivity. This may be true, but then the research questions have to be adjusted accordingly. What I mean is that scientific inquiry cannot 'have its cake and eat it too'. We can suffice in recording word lists or assembling data from television and the internet. This affords us the social distance that a more positivistic approach warrants, and many researchers do excellent work with this sort of material. If, however, we want to understand the finer details of vernacular events – events *defined* by the intimacy between interlocutors – then we have to get intimate. If we are unwilling or unable to do that, then we simply do not get to ask those questions. This is about the entitlement I have as a researcher to my humanity and the entitlements these participants have to their humanity.

Nonetheless, the intimacies of ethnographic work can be a messy affair. Eckert (1989a) offers one such example.

> But we frequently do not know, once we have gone away, the kinds of regrets they may have for having told us these things. During my work in the Pyrenees, I once had occasion to interview a normally quite reticent person, who 'opened up' during the interview and left me feeling rewarded and fulfilled. She, however, suffered a week of sleepless nights worrying about her indiscretion. As far I could see, she had not been indiscreet; she had not told me anything that was not common knowledge. But to this day, some 15 years later, I believe that she

feels a tug of regret every time she sees me (Eckert 1989a:34–35).

There have certainly been similar moments in my interviews. I have had a participants confess all sorts of private sentiments to me. How will they feel if, after four years, we suddenly happen across one another? Or worse, how will they feel if, after some years, I end up working with some colleague or ex-lover whom they bad-mouthed? Any and all of these scenarios are possible – especially in an isolated, close-knit city like Stockholm. Some of the participants may eventually also read this dissertation, so the best I can do is speak directly to them here and offer a second reassurance. No matter what, your secret is safe with me. My personal motive is to tell *Stockholm's* story and give back to a city that gave me so much when I had so little.

3.2 Consent and compensation

I told the participants that I was a doctoral researcher conducting research on the daily life and speech of today's Stockholmers. The participants were given an information sheet in Swedish that told them that the project was about daily life and language in Stockholm along with a four-tiered consent form. Copies of both are in Appendix A. The first tier of consent secured their participation. The second tier asserted that they had read the information sheet. The third tier asserted that they understood that I would appropriately anonymize their data in compliance with British and Swedish law. The fourth tier requested their permission to play their voices aloud at academic conferences or seminars. All 36 participants consented to every tier.

Once the interview was complete, I provided additional details about my interest in how social class, ethnicity, and contextual style varies depending on who is speaking.

Participants received 100 Swedish kronor (8£) per session, which is the typical rate for Humanities research in Sweden. Funding was provided by the British Economic and Social Research Council (ESRC).

4 Funding

This project was funded by the British Economic and Social Research Council (ESRC) by means of a three-year doctoral maintenance stipend administered by the London Social Sciences Doctoral Training Centre (approximate monetary value: £100 200). Supplementary funding was also provided by the ESRC by means of an Overseas Fieldwork grant (£450) and an Overseas Institutional Visit grant (£4 500).

The School of Languages Linguistics and Film at Queen Mary, University of London, also contributed financially to the project by means of its Research Training and Support Grant (£2 000) and three annual Postgraduate Research Funding grants (total: £3 000).

Additional funding (40 000 SEK ≈ £3 400) was also provided by the Sven och Dagmar Saléns stiftelse for transcription expenses.

5 Recording and transcription

Recordings were made on individual Zoom H1 recorders with self-powered Audio-Technica lavalier microphones. They are in WAV format, mono, with a sample rate of 16 000 Hertz. Back-up recordings were made by means of Zoom H4 recorders placed in proximity to the participant. No errors occurred with any of the main devices, so the back-up recordings were deleted.

Transcription was made on the ELAN software suite (Sloetjes & Wittenburg 2008). I broke up all recordings into breath groups, and I hired three transcribers to enter text into the ELAN cells (with the exception of sensitive material, which I transcribed myself). The transcribed material was then proofed and edited by me. The transcribers are Stockholm-based and were recruited, interviewed, and trained by me. They were contracted and compensated in compliance with applicable Swedish employment and tax statute.

The transcriptions were made according to an adapted version of the Philadelphia Neighborhood Corpus (2003) guideline, which itself is an adaptation of the transcription guideline for the SLX Corpus of Classic Sociolinguistic Interviews (Linguistic Data Consortium, 2003).

6 Phonetic segmentation: Constructing the SweFA forced aligner

Big data is necessary for any investigation that wishes to credibly generalize its findings beyond its participant sample to the greater speech community. Automatic segmentation or *forced alignment* is necessary for any phonetic project that analyzes 'big data'. A substantial portion of this project, nearly the entirety of 2017, was devoted to building a forced aligner for Swedish. This occurred in the form of three sub-projects: (1) I built a prototype from FAVE (Forced Alignment and Vowel Extraction) to align an initial batch of files; (2) I manually corrected the alignments in this first batch and used the errors from that batch to develop a pronunciation dictionary for vernacular Swedish; (3) I built a work-around for the Montreal Forced Aligner so I could use manually-corrected alignments to train it for Swedish. I will describe this process below.

6.1 Introducing forced alignment

Four forced-alignment programs that circulate widely for the English language are *Prosody-Lab aligner* (Gorman, Howell, & Wagner 2011) and its newer sister *The Montreal Forced Aligner* (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger 2017), *LaBB-CAT Transcriber* (Fromont & Hay 2012), and *Forced Alignment and Vowel Extraction* (*FAVE*, Rosenfelder, Fruehwald, Evanini, and Yuan 2011). Prosodylab aligner and the Montreal Forced Aligner market themselves as easily trainable for other languages. A fifth option is the *BAS Speech Science Web Services* (Kisler et al. 2016), which offers *WebMAUS Basic* for automatic transcription of Basque, Catalan, Dutch, English, Estonian, Finnish, Georgian, German, Japanese, Hungarian, Italian, Maltese, Polish, Russian, and Spanish.

For all five programs, the inputs are (1) an orthographic transcription, (2) a sound file, and (3) a pronunciation dictionary. The output is a phonetically-segmented file. The orthographic transcription is typically a tab-delimited file produced by ELAN (Sloetjes & Wittenburg 2008) that has start and end times for each breath group (see Figure 4.3). The pronunciation dictionary is a text file that has all possible pronunciation entries for every word in the language, which often can be as high as 30 or 40 possibilities for long compound words (see Figure 4.4) The output is a Textgrid file for use with Praat (see Figure 4.5, Boersma and Weenink 2017).

All of the programs are free of cost, and they provide various amounts of source code to the public along with varying degrees of written instructions for customizing the software to new languages. FAVE stands out because it was specifically designed for sociolinguistic purposes and because it has shown the highest accuracy rates for the alignment of vernacular speech (Yuan et al. 2013). Prosodylab and the Montreal Forced Aligner stand out because they provide the most robust assistance for training new languages.

FAVE and the Montreal Forced Aligned also stand out because they are designed to handle large sound files. They break files into chunks, transcribe them, and concatenate them back together – all behind the scenes. This is key for any large-scale sociolinguistic project such as mine, where sound files are up to three hours long. The other three programs require the user to manually break sound and transcription files down into one file per breath group, which would be logistically unfeasible for this project.

Accurate boundary alignment is not the biggest time-saver

One general misconception about automatic segmentation tools is that more accurate boundary placements always save more time. This is really only true if one wishes to extract data from uncorrected files, which is often the practice for projects that take formant measurements from the nucleus of, for example, 25 000 vowels within a corpus (Dodsworth & Benton 2017:377). However, for analyses of rhythm, manual corrections are required, and the time needed to manually move an incorrect boundary is the same for 0.002 seconds off-mark

```
August   0.0     2.373  Cirkusen var på väg! Deras plakat,
August   2.373   4.401  med bilden av en flygande cirkusartist
August   4.401   6.114  var uppsatta över hela stan.
August   6.114   9.426  Tidigare på dagen satt jag uttråkad när
                        morfar ringde åsa:
August   9.426  11.499  Jag har en överraskning till dig
August  11.499  13.862  något bra för humöret åsjälen
August  13.862  15.611  Vill du veta vad de är?
August  15.611  19.015  Ja! Berätta, berätta, berätta!
August  19.015  20.462  ropade jag förtjust.
August  20.462  22.302  Du får se själv i kväll.
August  22.302  23.797  Vad kul!
```

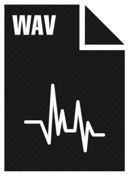


Figure 4.3: INPUTS 1 and 2
Five-column tab-delimited transcription input for FAVE, produced with ELAN, and sound file.

```
AV AA1 V
BILDEN  B IH1 LL D EH0
BILDEN  B IH1 LL D EH0 N
CIRKUSARTIST  S IH5  KK RS  RT IH6 SS
CIRKUSARTIST  S IH5  KK RS  RT IH6 SS T
CIRKUSARTIST  S IH5  KK S  RT IH6 SS
CIRKUSARTIST  S IH5  KK S  RT IH6 SS T
CIRKUSARTIST  S IH5  KK UH0 RS RT IH6 SS
CIRKUSARTIST  S IH5  KK UH0 RS RT IH6 SS T
CIRKUSARTIST  S IH5  KK UH0 S AH0  RT IH6 SS
CIRKUSARTIST  S IH5  KK UH0 S AH0  RT IH6 SS T
CIRKUSARTIST  S IH5  KK UH0 S RT IH6 SS
CIRKUSARTIST  S IH5  KK UH0 S RT IH6 SS T
CIRKUSARTIST  S IH5 RR K RS  RT IH6 SS
CIRKUSARTIST  S IH5 RR K RS  RT IH6 SS T
CIRKUSARTIST  S IH5 RR K S  RT IH6 SS
CIRKUSARTIST  S IH5 RR K S  RT IH6 SS T
CIRKUSARTIST  S IH5 RR K UH0 RS RT IH6 SS
CIRKUSARTIST  S IH5 RR K UH0 RS RT IH6 SS T
CIRKUSARTIST  S IH5 RR K UH0 S AH0  RT IH6 SS
CIRKUSARTIST  S IH5 RR K UH0 S AH0  RT IH6 SS T
CIRKUSARTIST  S IH5 RR K UH0 S RT IH6 SS
CIRKUSARTIST  S IH5 RR K UH0 S RT IH6 SS T
CIRKUSEN S IH1 KK UH0 S EH0
CIRKUSEN S IH1 KK UH0 S EH0 N
CIRKUSEN S IH1 RR K UH0 S EH0
CIRKUSEN S IH1 RR K UH0 S EH0 N
DERAS D EE3 R AH4 S
EN AEH1 NN
EN EE1 N
FLYGANDE F L YY3 G AH4 N EH0
FLYGANDE F L YY3 G AH4 D EH0
FLYGANDE F L YY3 G AH4 N D EH0
MED   M AEH1
MED   M EE1 D
PÅ_VÄG  P OAH0 V AE1 G
PÅ_VÄG  P V AE1 G
PLAKAT  P L AH0 K AA1 T
VAR   V AA1
VAR   V AA1 R
```

Figure 4.4: INPUT 3
Pronunciation dictionary with all possible pronunciations using ASCII characters for IPA.

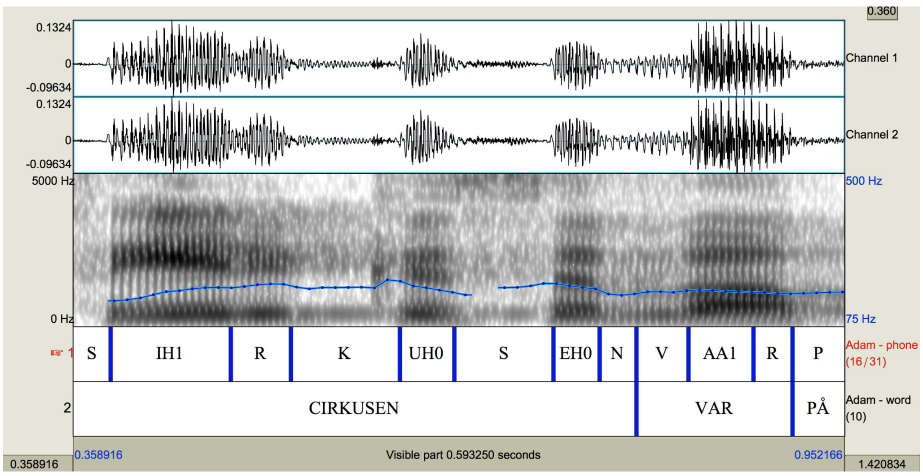


Figure 4.5: OUTPUT
Phonetically segmented file that is readable in Praat

as it is for 0.200 seconds off-mark. What saves time is fewer inaccurate boundary placements, with the degree of accuracy being more or less unimportant once the boundary error crosses about 0.007 seconds.

And those time savings are marginal when compared to the time needed to manually build boundaries and populate the resulting cells with the appropriate pronunciation orthography. To illustrate what I mean, I provide the following example. The recording that contains the first breath group from Figure 4.3 "cirkusen var på väg. Deras plakat"[2] lasts 2.37 seconds. The following tasks are performed to phonetically segment it manually with the time for each next to the task:

1. Building boundaries between words; populating the resulting cells: 2m 26s
2. Building boundaries between phonemes; populating the resulting cells: 4m 24s
3. Proofing boundaries; making final edits: 2m 19s

It takes 9 minutes 9 seconds (549 seconds) to manually align a 2.37-second transcription, which makes my recording-to-segmentation ratio 231:1[3]. Notice, however, that more than 75% of that time is spent building the boundary architecture and populating cells with the appropriate grapheme. Any program that can automatically do that is going to save a tremendous amount of time, regardless of how accurate boundary placement is. A program that can accurately place the boundaries is also a huge boon, but that is a secondary challenge.

Fortunately, no matter what system one chooses for alignment, tasks 1 and 2 will be created accurately if the pronunciation dictionary used for the automatic segmentation program is comprehensive. This is because the pronunciation dictionary determines how many boundaries to build before the breath group is analyzed. So when the user inputs the orthographic transcription "cirkusen var på väg. Deras plakat", the software first scans the following transcription text file.

```
August August  0.0 2.373  Cirkusen var på väg. Deras plakat
```

It then pulls up each word in the transcription via a pronunciation dictionary like below.

```
CIRKUSEN S IH1 KK UH0 S EH0
CIRKUSEN S IH1 KK UH0 S EH0 N
CIRKUSEN S IH1 RR K UH0 S EH0
CIRKUSEN S IH1 RR K UH0 S EH0 N
DERAS  D EE1  R AH0 S
PÅ_VÄG  P OAH0 V AE1 G
PÅ_VÄG  P V AE1 G
PLAKAT P L AH0 K AA1 T
VAR  V AA1 R
```

Figure 4.6: Miniature pronunciation dictionary

There are nine strings in Figure 4.6 The first space in each string tells the program that the orthographic lookup has ended and that it should begin reading for phonemes. Subsequent

[2] Translation: "The circus was on its way. Their poster"
[3] 400:1 is reported to be the typical upper limit (Yuan, Ryant, Liberman, Stolcke, Mitra & Wang, 2013).

spaces are interpreted as boundaries between phonemes. They are then immediately programmed into the Textgrid file as cells with unspecified boundaries.

Where the various segmentation programs are differentiated is in how they determine *where* the boundaries should go. That is determined by machine learning, which formulates a series of probabilities for where to place the boundaries based on prior inputs once the above empty architecture has been determined from the pronunciation dictionary. FAVE, LaBB-CAT, Prosodylab aligner and WebMAUS all use something called the *Hidden Markov Toolkit* (HTK, Young, Woodland, and Byrne 1993). The Montreal Forced Aligner uses Kaldi (Povey et al. 2011).

For situations where there are multiple pronunciations, as in CIRKUSEN and PÅ_VÄG, the program uses the first lookup first, runs boundary-location assessments, and then loops to the next option and so on (and so on and so on) for as many options as there are. It then assesses the best fit.

With the above in mind, the smallest piece was building a prototype from FAVE, and the largest piece was building a comprehensive pronunciation dictionary that includes all possible vernacular pronunciations. Once those two pieces were safely in place, I constructed a workaround that allowed me to train manually-corrected files on the Montreal Forced Aligner.

6.2 Building a forced aligner for Swedish

Stage 1: The Prototype from FAVE

FAVE has transparent architecture and ample documentation, which makes it particularly apt for prototype adaptation. I therefore went into its code and relabeled – and duplicated when necessary – its hidden Markov models from American English to the closest corresponding Swedish phoneme. The English monophones in FAVE are taken from the Advanced Research Projects Agency's (ARPA) Speech Understanding Project and are entitled *ARPAbet*. I created a parallel system for Swedish entitled *SweFAbet*. Table 4.2 provides a full list of the Swedish phoneme inventory. The first column contains the SweFAbet monophone, the second column the corresponding IPA symbol, the third column the most common corresponding grapheme, the fourth column a Swedish lexical example (some are loanwords), the fifth column the closest English phoneme, and the sixth column the ARPAbet monophone for that closest English phoneme.

After I programmed these substitutions in, I also built a prototype for Danish (not used for this project). With the help of Michael McGarrah at Georgia Institute of Technology, I built a unicode-8 converter because HTK is only ASCII-compatible. McGarrah then built a language-general shell to host the Danish, English, and Swedish aligners within one program (Young & McGarrah 2017). I subsequently turned my focus to procuring and adapting a pronunciation dictionary.

Consonants

SweFAbet	IPA	grapheme	lexical example	closest English phoneme	ARPAbet
P	p	p	pil	p	P
PP	pː	pp	trappa	p	P
B	b	b	bil	b	B
BB	bː	bb	svabba	b	B
T	t	t	tal	t	T
TT	tː	tt	fatta	t	T
D	d	d	dal	d	D
DD	dː	dd	ladda	d	D
K	k	k	kal	k	K
KK	kː	ck	macka	k	K
G	g	g	gas	g	G
GG	gː	gg	tagga	g	G
M	m	m	mil	m	M
N	n	n	nål	n	N
NG	ŋ	ng, gn	ring, lugn	ŋ	NG
R	r	r	ris	ɾ	D
F	f	f	fil	f	R
V	v	v	vår	v	V
TH	θ	th	thriller	th	TH
DH	ð	th	the	th	DH
S	s	s	sil	s	S
Z	z	z	guzz	z	Z
TJ	(C)	tj	tjock	ʃ	SH
SJ	ɧ	sj, sk, stj	sjuk	h	HH
HH	h	h	hal	h	HH
J	ʝ	j	jag	j	Y
L	l	l	lös	l	L
JH	dʒ	g, j	Jaffar	dʒ	JH
W	w	w	walla	w	W
CH	tʃ	c, ch	çok	tʃ	CH
RT	ʈ	rt	fart	t	T
RD	ɖ	rd	bord	d	D
RN	ɳ	rn	barn	n	N
RS	ʂ	rs	fors	ʃ	SH
RL	ɭ	rl	Karl	l	L

Vowels

SweFAbet	IPA	grapheme	LEXICAL SET / example	closest English phoneme	ARPAbet
II	iː	i	DIS	i	IY
YY	yː	y	TYP	i	IY
UU	ʉː	u	LUS	u	UW
EE	eː	e	LETA	i	IY
OE	øː	ö	SÖT	ʊ	UH
OEE	œː	ö(+r)	DÖR	ʊ	UH
AE	ɛː	ä	NÄT	ɛ	EH
AEE	æː	ä(+r)	LÄR	æ	AE
OO	uː	o	SOT	u	UW
OA	oː	å	LÅS	oʊ	OW
AA	ɑː	a	LAT	ɔ	AO
IH	ɪ	i	DISK	ɪ	IH
YH	ʏ	y	FLYTTA	ɪ	IH
EH	ɛ	e	LETT	ɛ	EH
OEH	œ	ö	DÖRR	ʊ	UH
AEH	ɛ	ä	SÄRK	ɛ	EH
OH	ʊ	o	ROTT	ʊ	UH
UH	ɵ	u	LUDD	ʊ	UH
OAH	ɔ	å	LOTT	ɔ	AO
AH	a	a	LASS	a	AA
AJ	aʝ	aj	Eileen	aɪ	AY
OJ	oʝ	oj	Freudenthal	ɔɪ	OY
EJ	ɛʝ	ej	Heijdenberg	eɪ	EY
EU	ɛ͡u	eu	Europa	ɛ	E
AU	a͡u	au	Gaude	aʊ	AW

Table 4.2: SweFAbet, corresponding IPA, grapheme, lexical example, and closest English phoneme with ARPAbet

Stage 2: The pronunciation dictionary

As demonstrated in the prior section, 75% of the time saved in automatic transcription comes from having a pronunciation dictionary that is both comprehensive and accurate. The dictionary must be formatted with the orthographic entry first, followed by a space followed by the monophones demarcated with a space between them.

For Swedish, two resources were particularly rich:

1. *Folkets Lexikon*: This is an open-source dictionary created and managed by the School of Computer Science and Communication at the KTH Royal Institute of Technology in Stockholm, Sweden.

2. *The NTS pronunciation dictionary* from Nordic Language Technology Holdings, Inc.

```
<word value="plakat" lang="sv" class="nn"><translation value="placard" />
<translation value="poster" />
<phonetic value="plakA:t" soundFile="plakat.swf" />
<paradigm><inflection value="plakatet" />
<inflection value="plakat" />
<inflection value="plakaten" />
</paradigm>
<synonym value="affisch" level="4.0" />
<synonym value="berusad" level="3.3" />
<synonym value="full" level="3.0" />
<synonym value="poster" level="3.1" />
<synonym value="skitfull" level="4.0" />
<synonym value="skylt" level="4.5" />
<see value="plakat||plakat..1||plakat..nn.1" type="saldo" />
<example value="ett plakat med stora röda bokstäver"><translation value="a poster (
    → placard) with large red letters" />
</example>
<compound value="valplakat"><translation value="election poster (placard)" />
</compound>
<definition value="stort anslag" />
</word>
```

(a) Before conversion

```
plakat plakA:t
plakatet
plakat
plakaten
```

(b) After conversion

Figure 4.7: Entry for the word 'plakat' in Folkets Lexikon

Folkets Lexikon Most proprietary dictionaries of Swedish are actually interface improvements to the *Folkets Lexikon* (*People's Lexicon*, Kann 2010; Kann and Hollman 2011). This comprehensive word list was a government-funded project to provide a comprehensive wordlist on the web. It was first published in 2009 and has undergone successive improvements through 2014. Its XDXF format is illustrated in Figure 4.7(a). Using *BBEdit* for Mac, I wrote a series of regular expressions to transform it into a usable format. The finished product is shown in Figure 4.7(b).

Note that in Figure 4.7 the root form offers a phonetic pronunciation, but the inflected forms do not. In fact, Folkets Lexikon, while claiming to have 200 000 total entries, only has 18 928 pronunciation entries. This was insufficient for a project on vernacular speech.

NTS pronunciation dictionary for Swedish In 2003, Nordic Language Technology Holdings, Inc. (NTS AB) went bankrupt, and the local council seized its intellectual property and handed it over to Norway's National Library. At the time, NTS was working on Automatic Speech Recognition (ASR) and production for Danish, Norwegian bokmål, and Swedish. All three pronunciation dictionaries have been made publicly available.

The dictionaries are formatted in code unique to the IBM ASR software that they were customizing, so I wrote a series of regular expressions to get it into the SweFA prototype format. Figure 4.8(a) illustrates the original code, and Figure 4.8(b) illustrates my conversion. The resulting product was a pronunciation dictionary with 927 167 entries. I reconciled its entries with Folkets Lexikon to test its scope. The NTS entries matched all but 609 of the 18 928 Folkets Lexikon entries. I then added approximately 3 000 slang words from the web-

```
plane;JJ;SIN|DEF|NOM|MAS|POS;plane;;INFL;SWE;;;;;""plA:$ne;0;STD;SWE;;;;;;;;;;;;;583845;
      → inflector;;INFLECTED;plan|75878;a1q-vrigt;a1;;;;;;;;;;;;;plane;;;1583845
planekonomi;NN;SIN|IND|NOM|UTR;plan+ekonomi;NN+NN;LEX|INFL;SWE;;;;;""plA:n$e$kU$nU$%mi:;1;
      → STD;SWE;;;;;;;;;;;;;36259;enter_se|inflector;;INFLECTED;planekonomi|75883;s3s-k;s3
      → ;;;;;;;;;;;;;planekonomi;;;73176
planekonomier;NN;PLU|IND|NOM|UTR;plan+ekonomier;;INFL;SWE;;;;;""plA:n$e$kU$nU$%mi:$er;0;STD;
      → SWE;;;;;;;;;;;;;583895;inflector;;INFLECTED;planekonomi|75883;s3s-k;s3;;;;;;;;;;;;;
      → planekonomier;;;1583895
planekonomierna;NN;PLU|DEF|NOM|UTR;plan+ekonomierna;;INFL;SWE;;;;;""plA:n$e$kU$nU$%mi:$e$n'a
      → ;0;STD;SWE;;;;;;;;;;;;;583896;inflector;;INFLECTED;planekonomi|75883;s3s-k;s3
      → ;;;;;;;;;;;;;planekonomierna;;;1583896
planekonomiernas;NN;PLU|DEF|GEN|UTR;plan+ekonomiernas;;INFL;SWE;;;;;""plA:n$e$kU$nU$%mi:$e$n
      → 'as;0;STD;SWE;;;;;;;;;;;;;583900;inflector;;INFLECTED;planekonomi|75883;s3s-k;s3
      → ;;;;;;;;;;;;;planekonomiernas;;;1583900
planekonomiers;NN;PLU|IND|GEN|UTR;plan+ekonomiers;;INFL;SWE;;;;;""plA:n$e$kU$nU$%mi:$es';0;
      → STD;SWE;;;;;;;;;;;;;583899;inflector;;INFLECTED;planekonomi|75883;s3s-k;s3
      → ;;;;;;;;;;;;;planekonomiers;;;1583899
planekonomin;NN;SIN|DEF|NOM|UTR;plan+ekonomin;;LEX|INFL;SWE;;;;;""plA:n$e$kU$nU$%mi:n;1;STD;
      → SWE;;;;;;;;;;;;;58434;enter_se|inflector;;INFLECTED;planekonomi|75883;s3s-k;s3
      → ;;;;;;;;;;;;;planekonomin;;;73177
planekonomins;NN;SIN|DEF|GEN|UTR;plan+ekonomins;;INFL;SWE;;;;;""plA:n$e$kU$nU$%mi:ns;0;STD;
      → SWE;;;;;;;;;;;;;583898;inflector;;INFLECTED;planekonomi|75883;s3s-k;s3;;;;;;;;;;;;;
      → planekonomins;;;1583898
planekonomis;NN;SIN|IND|GEN|UTR;plan+ekonomis;;INFL;SWE;;;;;""plA:n$e$kU$nU$%mi:s;0;STD;SWE
      → ;;;;;;;;;;;;;;583897;inflector;;INFLECTED;planekonomi|75883;s3s-k;s3;;;;;;;;;;;;;
      → planekonomis;;;1583897
planekonomiska;JJ;|||UTR-NEU;plan+ekonomiska;;LEX;SWE;;;;;""plA:n$e$kU$%no:$mI$ska;1;STD;SWE
      → ;;;;;;;;;;;;;;;dictatraining_se;;;;;;;;;;;;;;;;;;planekonomiska;;;127666
planen;NN;PLU|DEF|NOM|NEU;planen;;INFL;SWE;;;;;"plA:$nen;0;STD;SWE;;;;;;;;;;;;;583868;
      → inflector;;INFLECTED;plan|75880;s7n-vriga ex trd;s7;;;;;;;;;;;;;planen;;;1583868
```

(a) Before conversion

```
PLANE  P L AA3 N EH4
PLANEKONOMI P L AA5 N EHO K OHO  N OHO  M II6
PLANEKONOMIER P L AA5 N EHO K OHO N  M II6 EH7 R
PLANEKONOMIERNA P L AA5 N EHO K OHO N OHO  M II6 EH7  RN AHO
PLANEKONOMIERNAS P L AA5 N EHO K OHO N OHO  M II6 EH7  RN AHO S
PLANEKONOMIERS P L AA5 N EHO K OHO N OHO  M II6 EH7 RS
PLANEKONOMIN P L AA5 N EHO K OHO  N OHO  M II6 N
PLANEKONOMINS P L AA5 N EHO K OHO  N OHO  M II6 N S
PLANEKONOMIS P L AA5 N EHO K OHO  N OHO  M II6 S
PLANEKONOMISKA P L AA5 N EHO K OHO N OA6  M IH7  S K AHO
PLANEN  P L AA1  N EHO N
```

(b) After conversion

Figure 4.8: Snapshot of entries from NTS word list and their subsequent SweFA transformations

site slangopedia.se and manually entered in pronunciations. Most of the words were familiar to me, so I coded in pronunciations based on my own knowledge. Where I was unsure, I sought advice from native Swedes in my own personal network.

The final pronunciation dictionary was used to align small batches of sound files. I would then manually correct the sound files and document any pronunciation differences. These differences would then be programmed into the dictionary for all words where they were relevant.

For example, in one casual recording, the speaker pronounced the word *presenˈterade* like *pznˈtera*. I made a note of this during my manual corrections and then wrote a series of regular expressions that would add a *pz* pronunciation for all occurrences of unstressed *pres*, a *sn* pronunciation for all occurrences of unstressed *sen*, and the elision of word-final *de* after unstressed *a*.:

```
PRESENTERADE P R EHO S EHO N T EE1 R AHO
PRESENTERADE P R EHO S EHO N T EE1 R AHO D EHO
PRESENTERADE P R EHO S EHO T EE1 R AHO
PRESENTERADE P R EHO S EHO T EE1 R AHO D EHO
PRESENTERADE P R EHO S N T EE1 R AHO
PRESENTERADE P R EHO S N T EE1 R AHO D EHO
```

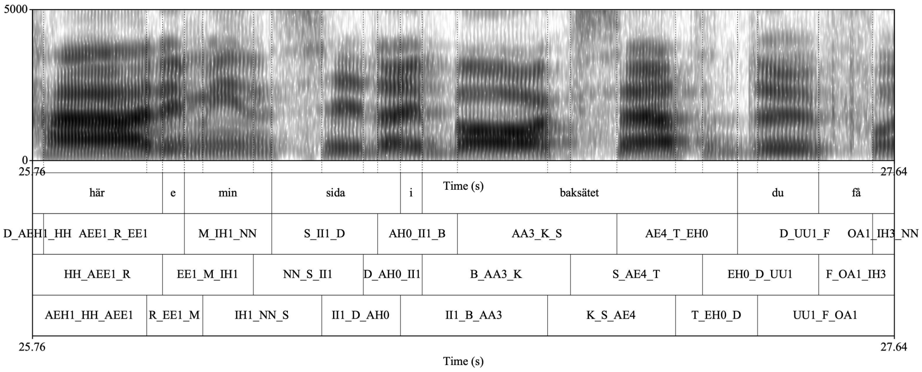


Figure 4.9: Fictitious triphone words that I developed from the manually-corrected corpus.

```
PRESENTERADE P Z EH0 N T EE1 R AH0
PRESENTERADE P Z EH0 N T EE1 R AH0 D EH0
PRESENTERADE P Z EH0 T EE1 R AH0
PRESENTERADE P Z EH0 T EE1 R AH0 D EH0
PRESENTERADE P Z N T EE1 R AH0
PRESENTERADE P Z N T EE1 R AH0 D EH0
```

After approximately 6-9 months of bootstrapping, I had aligned and manually corrected approximately 30 percent of the corpus, and the dictionary had begun to cover nearly all pronunciations. The final version has just over 16 million entries.

Stage 3: Training the Montreal Forced Aligner

With 30 percent of the corpus aligned, I began training the Montreal Forced Aligner (MFA) on the manually-corrected data. This is not a straightforward process because MFA does not typically train in that manner. MFA, like all other forced aligners, is designed to train at the utterance / breath-group level. Referring back to Figure 4.3, this means that the boundaries for the phonemes in those utterances are modeled by assessing what those utterances have in common with each other based on the pronunciation dictionary entries. I performed this task on my corpus at the beginning and the resulting product was poor.

This is complicated by the fact that individual phonemes cannot be entered into the training model because there is a minimum 100-millisecond requirement for all utterance entries. The workaround I came up with was to derive fictitious triphone entries from the manually-corrected corpus. This is demonstrated in Figure 4.9. Recall that the final pronunciation dictionary would have multiple pronunciation entries for each word lookup on tier 1, especially *här*, *min*, *sida*, and *baksätet*. This would add unnecessary noise to the training. Therefore, instead of training it on the words, I constructed fictitious triphone Praat Textgrids that corresponded to every corrected boundary in the corpus, shown on tiers 2–4. These triphones were then added to a fictitious pronunciation dictionary as individual words with single pronunciation options. I then trained the aligner on the fictitious dictionary and the fictitious Textgrids.

The remainder of the corpus was aligned with the above method in a bootstrapping manner. As I corrected more files, updated training was run. More manual corrections were made on the updated batch and added to the training. I continued this cycle until the entire corpus had been aligned and manually corrected.

7 Theoretical approaches to racialization and class

The analysis chapters rely heavily on the operationalization of two key predictors: *racialization* and *social class*. These, however, are controversial constructions that first must be problematized before any analysis can proceed. The next two subsections therefore are devoted to an explanation of how I understand racialization and class in the context of Stockholm, Sweden, and Scandinavia.

7.1 *Racialization* in Sweden

I use the terms *svensk* and *invandrare* in this thesis. The words translate literally as 'Swedish' and 'immigrant', respectively. Although the terms ostensibly refer to properties of national and migrant origin, they are in fact used in Sweden as terms of race and racialization (Hübinette, Hörnfeldt, Farahani, and Rosales 2012,Young 2018c:43). Throughout this thesis, the terms are used to refer to their standard Swedish connotations of racialization instead of actual national or migration background.

Ras (Eng. *race*) is a controversial term in Sweden. Its mention will elicit derision and perhaps also a declaration that Sweden has long moved past its painful role in the promotion of race biology[4]. In the glib words of Molina and De los Reyes (2002), "it is difficult and just plain unacceptable to think that such notions of race can occur more than as mere exceptions in the Swedish welfare state, a modernist project minted of equality and humanism[i]" (2002:297). Sweden also takes a fundamentally different approach to race at an official level than the UK and the US. The Swedish state acknowledges neither race or ethnicity and solely collects data on country of birth in its census; the UK and US collect data on both ethnicity and race. These differences materialize often in the general population's meta-discourse.

> In America today there is a proposed bill in Congress to add a fifth category to the four current racial categories: 'multiracial'. This illuminates the hopeless and absurd use of an archaic term[ii]. (Werbart 2000:183)

Despite metapragmatic derision to race as a notion, the assignment of *invandrare* and *svensk* is nearly always phenotypic. This is especially the case if both subjects speak fluent Swedish. Naturally, an accented white speaker will not be called *svensk*, but an unaccented

[4] Sweden's State Institute for Racial Biology in Uppsala was one of the world's more prominent institutions for eugenics and race biology in the early Twentieth Century.

non-white speaker will be called *invandrare*. I offer an anecdotal example. Before going into academia, I worked with two women in the nightlife industry who had Swedish mothers. One had a Gambian father, and the other had a white American father. Both spoke classic Stockholmian with some *södersnack* features. The former woman was often referred to by customers and colleagues as *invandrare* or *invandrare bakgrund*[5] (immigrant background) and the latter woman as *svensk*.

This example draws attention to the fact that the terms are phenotype-based appellations. *Phenotype* refers specifically to the consensus in Sweden on what constellations of physical features 'look Swedish' and what do not. Like most constructions, the indexical threshold between *svensk* and *invandrare* depends on a some sort of complicated weighted hierarchy. In simple terms, someone with a certain nose might pass as *svensk* if his/her hair is blond while someone with the exact same nose might not with brown hair. And despite the obvious fact that this is a heavily manufactured phenomenon, it organizes contemporary Swedish society in a very real and material way. Hübinette et al. (2012) provide examples of these material consequences.

> [Swedish whiteness] stands out per one unique facet. It is the exclusivity associated with Swedish whiteness that makes it indubitably more privileged than the whiteness found in other Western nation states. This hallmark of Swedish whiteness has not just made it harder to be considered white in Sweden than in other Western countries, it has also resulted in an indigenous hierarchy under which various degrees of whiteness have sprouted. This Swedish notion of whiteness manifests itself in such a way that members of the *majority* population who are short or have dark hair, dark eyes, or darker skin also can be subjected to racialization, be mistaken for *invandrare* (immigrant), and draw questions about where they come from and why they look 'un-Swedish'. [...] [M]embers from both the majority and minority populations in Sweden even bleach their hair, take growth hormones, and subject themselves to medical procedures in order to achieve the idealized Swedish body norm.[iii] (Hübinette et al. 2012:60)

The aforementioned point is important to the study of multiethnolects on a number of fronts. Researchers from outside of Scandinavia might marvel at the "high number of white multiethnolect speakers" without realizing that the speakers, in fact, are not 'white' at all. A large portion of multiethnolect speakers in Stockholm are of Greek and South Slavic descent. These speakers would be considered 'white' in the UK or the US, but are typically racialized in Sweden. In Denmark, a country with a nearly identical racialized binary (*dansk/indvandrer*), a recent television program called 'Paki Danish' (*Perkerdansk*) drew

[5]The media will also sometimes use 'svensk bakgrund' (Swedish background) for white subjects and 'invandrare bakgrund' (immigrant background) to refer to subjects who appear phenotypically different. Actual investigations into whether the white subjects might have parents from, say, other European countries are never conducted; nor are investigations into whether the non-white subjects might have Swedish-born parents.

'Race'	Ethnicity	'Race'	Ethnicity
invandrare	Chilean	svensk	Chilean / Swedish
invandrare	Egyptian Arab	svensk	Finland Swedish*
invandrare	Eritrean	svensk	Finnish / Swedish
invandrare	Greek / Azeri	svensk	Norwegian / Swedish
invandrare	Iranian	svensk	Norwegian / Swedish
invandrare	Iranian	svensk	Swedish Jew / Romanian Jew
invandrare	Iranian	svensk	Turkish / Swedish
invandrare	Kurd (Feyli)	svensk	Swedish
invandrare	Kurd (Kurmanji)	svensk	Swedish
invandrare	Kurd (Kurmanji)	svensk	Swedish
invandrare	Luo	svensk	Swedish
invandrare	Mbembe	svensk	Swedish
invandrare	Peruvian	svensk	Swedish
invandrare	Qashqai / Gilak	svensk	Swedish
invandrare	Somali	svensk	Swedish
invandrare	Tigrinyan	svensk	Swedish
invandrare	Tigrinyan	svensk	Swedish
invandrare	Tigrinyan		
invandrare	Turk		

* Finland Swedes are not to be confused with Finnish Swedes. The former are ethnic Swedes from the coastal regions of Finland where Swedish settlements have existed since the 12th century. The latter are ethnic Finns who have immigrated to Sweden.

Table 4.3: Breakdown of the ethnic groups that fall under the *svensk* and *invandrare* racialized binary.

attention to Denmark's similarly narrow definition of whiteness in relatively blunt terms: Bosnians speak 'Paki Danish', but can sometimes be visually mistaken for ethnic Danes.

> The typical Turk differentiates from ethnic Danes by having a slightly darker hue—both as it relates to skin and hair color. In Turkey it is very typical to have a lot of hair growth. [...] Iraqis are generally Caucasian in their appearance. [...] Bosnians have very light skin and can therefore be misidentified as ethnic Danes. Their eye color is on the lighter side of the scale. And actually a lot of Bosnians have blue or green eyes. [iv] (Danmarks Radio 3 2006)

As I alluded to in my anecdotal example above, the *svensk*/*invandrare* division has little to do with actual immigration status or legacy immigration status. In fact, I believe that a large portion of the majority population might have at least one parent or grandparent of non-Swedish ancestry, despite self-identifying as *helsvensk* ('whole Swedish'). There is little official data on this, but my participant pool indicates that this might be the case. Table 4.3 shows the breakdown of ethnicity for my participants for the racialized *svensk*/*invandrare* binary[6]. Of the 17 speakers who initially claimed to be *svensk*, seven of them have a parent from another country. One is even half Turkish but identifies as *helsvensk* (whole Swedish) due to the fact that he is 'light' (his own words) and was raised solely by his Swedish mother.

Molina and De los Reyes (2002) argue that the 'migrant issue' and 'questions of ethnic tension' have become distractions to the real issue at hand, namely a racialized population

[6] To avoid putting 'the cart before the horse', I do not offer a full oversight of the collected data until Chapter 6. However, this sliver of data is necessary to make my case for how I approach racialization.

born in Sweden yet not 'Swedish' enough. Neergaard (2017) supports this position and argues that hierarchies within Swedish trade unions cut across racial lines within the native-born population rather than actual immigrant status, further motivating a Critical Race perspective.

> [T]he term 'racialised workers' will be used to refer to those, with or without migrant background, who are subordinated within hierarchical divisions of labour, with ascribed stereotypical identities in terms of skin colour, culture, religion etc. This is based on assumed ethnic and racial differences, which, as negative attributes, affect their working conditions, professional mobility and public representation. (Neergaard 2017:87)

Cornips and de Rooij (2013) observe that in the Netherlands this process of racialization falls into a binary schema of those 'who belong' and those 'who do not'. Their data demonstrates that the schema is so powerful that everyone participates in its reinforcement:

> In Dutch society, the dominant classificatory schema allochthon/authochthon is wholly based on the idea of belonging to a specific place or not. The force of the Dutch dominant autochthon (from here) – allochthon (from elsewhere) binary schema may be so great that even those branded as allochthons may have internalized it and started applying it in processes of selfing and othering. (Cornips & de Rooij 2013:161)

In a later study, Cornips (2019, forthcoming) found that visible minorities who were born in Limburg and speak nativelike Limburgish are rarely viewed by strangers as legitimate speakers. When spoken to in standard Dutch, they have two choices. They can either reframe the speech event in Limburgish or they can respond in Dutch. A lifetime of encounters like this has resulted in some minority dialect speakers internalizing their interpellated illegitimacy, proclaiming, themselves, that minority dialect users sound 'weird' and 'stupid' (2019, forthcoming).

In similar fashion, the racializing system in Stockholm is also internalized by Swedes of visible-minority status to a degree that they typically will self-identify as *invandrare*[7] and position themselves as The Other[8] in relation to *svenskar*. Among the participants in this study, this is also the case, despite the fact that only three were actually born outside of Sweden (and even they all arrived before age seven).

7.2 Social class

I devise a detailed social-class index based on a Principle Components Analysis of six social factors. Because the calculation of this index has an analytic element to it, I have decided

[7] Hübinette et al. (2012) cites visible-minority Swedes who were adopted and raised by white Swedish parents as one group that actively resists its racialization. Nonetheless, for them this is an ongoing lifetime battle (2012:63-64).

[8] de Beauvoir (2009 [1949]:26–30)

to place it in the first analysis chapter (Chapter 6). The goal of this section, however, is to provide a theoretical framework and short literature review of social class.

Sweden is popularly thought of as a classless society. Within Swedish academia however, there exists wide acceptance that social class is a key framework by which Swedish society is organized. Stockholm is known to be a city where social class is particularly salient. It is common, in fact, for Swedes from other parts of the country to refer to Stockholmers as *dryg* (approximate translation: *conceited*; see, for example, Hellberg 2006).

Access to power shapes social practice

When power is distributed unevenly within a community, it shapes the social behavior of the agents involved. Linguistic practice cannot be abstracted from social behavior; the former is a subspecimen of the latter (Eckert & McConnell-Ginet 1992). According to Bourdieu (1987), *social class* is defined as groups of individuals who cluster via their dispositions and practices – dispositions and practices of which are formed by *shared access* to the *uneven rationing* of power.

> [C]onstructed classes can be characterized in a certain way as sets of agents who, by virtue of the fact that they occupy similar positions in social space (that is, in the distribution of powers), are subject to similar conditions of existence and conditioning factors and, as a result, are endowed with similar dispositions which prompt them to develop similar practices. (Bourdieu 1987:6)

One major way in which stratified access to capital (in its various forms) affects language use is by incubating stratified value systems. A generally-accepted generalization among sociolinguistics is that the value system of disadvantaged communities is governed by *solidarity* whereas the value system of privileged communities is governed by *status* (Brown and Gilman 1960; Milroy 1987; Trudgill 1971; Woolard 1985; among others). According to Woolard (1985), these competing 'arche-values' motivate 'illegitimate' language within the working classes and 'legitimate' language within the more privileged classes.

> Even where there is recognition of the authority of the legitimate language, there can be repudiation of its value on an important contrasting dimension. Competing sets of values exist, creating strong pressures in favor of the 'illegitimate' languages in the vernacular markets, and not just an absence of pressure against them. [...] It is nonetheless critical to understand that these vernacular practices are productive, not merely reproductive, that they arise not from a mere bending to the weight of authority, but are paradoxically a creative response to that authority, mediated by the oppositional value of solidarity. (Woolard 1985:744)

According to Labov (2001), the valorization of oppositional practices in disadvantaged communities – whether minority or working-class or both – incubates linguistic innovation. This process is known as the *Nonconformity Principle*.

> The Nonconformity Principle: Ongoing linguistic changes are emblematic of nonconformity to established social norms of appropriate behavior, and are generated in the social milieu that most consistently defies those norms. (Labov 2001:516)

The Nonconformity Principle was one of many theoretical paradigms that emerged from what Eckert (2012) calls the *First Wave*. Modeling a linguistic variable according to numerical representations of social class is a signature feature of First Wave studies in variationist sociolinguistics. These were systematic urban investigations of (mostly) phonetic variation that were conducted in the third quartile of the twentieth century (e.g., Fontanella de Weinberg 1974; Labov 2001, 2006 [1966]; Trudgill 1974; Wolfram 1969). The authors created numerical aggregates for social class and investigated the statistical correlation between the variation of social class with the variation of the linguistic variable. The numerical aggregate for social class was typically a mathematical mean. This mathematical mean was often calculated from numerical representations of artifacts of power: income, education level, father's education level, home ownership, and so on. Bourdieu (1980, 1987) refers to these artifacts of power as *capital*.

> In this respect, such classes meet all the requirements of a scientific taxonomy, at once predictive and descriptive, which allows us the greatest amount of information for the least cost: the categories obtained by cutting up sets characterized by the similarity of their occupational conditions within a three-dimensional space have a very high predictive capacity for a relatively small cognitive expense (that is, relatively little information is necessary to determine the position in that space: one needs three coordinates, global volume of capital, composition of capital, and social trajectory). (Bourdieu 1987:6)

The latter three coordinates highlight Bourdieu's fundamental framework for class. Agents possess varying amounts of capital, types of capital, and trajectories for mobility. These attributes cluster in meaningful ways while also forming hierarchies. For example, an electrician may earn the same salary as a university lecturer, representing a shared amount of *economic* capital. However, they both possess different degrees of *symbolic* capital, which can further divide into an unlimited number of subtypes. *Cultural* and *educational* capital are just some of these subtypes, and *linguistic* capital is yet another subtype. The very last coordinate, *trajectory*, seems to be particularly unique to modern Western urban communities. It is challenged by Rickford's (1986) findings in Guyana, something I discuss further along in this review.

According to Bourdieu (1980), some types of symbolic capital are more embodied than others. In his investigation of 1 216 French participants, Bourdieu found that educational level – i.e., formal qualifications – had the highest predictive power for cultural goods within overt domains like painters and musical works. In underspecified domains of taste, social background had high predictive power when educational level was held constant (1980:226). For example, most of the explanatory power for whether one preferred the *Blue Danube* (low-brow) over *Well-tempered Clavier* (high-brow) was formal educational credentials. This is because these pieces had circulated popularly for some time and had been imbued with overt cultural indexicality (1980:229). In contrast, social background (i.e., father's education) was a much stronger correlate for seemingly arbitrary aesthetic preferences. Participants from an upper-class background were more likely to view cabbages or a snake as 'beautiful' and a sunset as 'ugly' or 'banal' while participants from a working-class background were more likely to say the exact opposite. That is to say that aspects of social class are cognitively embedded. The less specified the social practice, the less control one has on its expression. The more specified the social practice, the more control one has.

Linguistic variation may be governed in similar ways. For example, one's formal education may correlate strongly with the degree to which one avoids overtly stigmatic features in speech (referred to by Labov 2001 as *markers*). One's childhood social background, however, may have more predictive power as to whether one can detect and avoid less obvious stigmatic features in language, including incipient classed features that few have yet detected (referred to by Labov 2001 as *indicators*). I discuss this in detail in Chapter 8 (p. 207).

Bourdieu's work focuses less on 'class', per se, and more on how various types and amounts of capital are embodied and translated into practice. In terms of this investigation, this is also how I view the notion of class. Rather than relying exclusively on a single social-class index, I choose also to examine how the granular components of class predict the distribution of rhythmic variation. The parameters I have selected for capturing overt and less cognitively-embedded forms of capital are income, occupational status, and formal education. The parameters I have selected for capturing covert and more cognitively-embedded forms of capital[9] are parental occupational status, parental formal education, and taste.

Social-class indices: Aggregating various forms of capital

The First Wave studies that I cite above used social-class indices that resemble those developed by Warner, Meeker, and Eells (1949) and Hollingshead and Redlich (1958), two of the first modern studies to use least-squares regression and factor analyses for the purpose of operationalizing social class.

Hollingshead and Redlich's (1958) Index of Social Position (ISP) was based on a semi-objective assessment of 552 families in New Haven, USA, who participated in interviews that

[9] Bourdieu uses the term *inculcation*, but I think it means the acquisition of cultural practices during the Critical Period in childhood when one possesses a high degree of cognitive plasticity. With the onset of puberty and beyond, plasticity reduces, and class practices become fossilized.

lasted between two and three hours. The interviewers placed each family into one of seven social classes according to their impression of the family's stylistic expression, and the authors then ran a regression analysis using seven-point scales for the following three factors: residence, education, and occupation. Residence had been established in earlier studies, enabling the authors to assign a status level to any address in New Haven. Education was enumerated according to institutional milestones in schooling (finished junior high school, high school, etc). Occupation was broken down according to the adjusted United States Census 'Alba Edwards' guidelines that offered a relatively crude hierarchy for professions. The resulting regression equation was $\text{ISP} = 0.185 \cdot \textit{Residence} + 0.154 \cdot \textit{Education} + 0.269 \cdot \textit{Occupation} + 0.884$. They discarded the constant and used the resulting coefficients for weights: 6, 5 and 9, respectively (1958:387–397). This index was then cross-validated with mass-media consumption. The hierarchical clustering of media choice accounted for 94% of the variation in ISP (1958:398–407).

Wolfram (1969:32–36) used a direct replication of Hollingshead and Redlich's scale in his sociolinguistic investigation of Black English in Detroit, USA. Similarly, Labov (2001) combined occupation, education, and residence in his study of the American city of Philadelphia. However, his index differed in two ways; residence was operationalized via home value, and the three components carried equal weights in the index.

Other sociolinguistic investigations developed their own indices (although many of them appear also to be inspired by Warner et al. and Hollingshead and Redlich). Labov (2006 [1966]) developed two indices for his investigation of New York. The first, which he called 'socioeconomic class' was established *a priori* in order to objectively model the variables (r), (æh), (oh), (th), and (dh). He assigned participants an ordinal numerical value for income, occupation, and education each that ranged between 0 and 3, rendering four possible levels for each of the three predictors. He then added them together, rendering a highest possible SEC of 9 and a lowest possible SEC of 0 (2006 [1966]:132-137). Gradient SEC offered weak predictive power for some of the binomially-distributed variables, so Labov (2006 [1966]) worked backwards to identify an index that appropriately correlated with the binomial split. He entitled this measurement 'social class' (SC) and constructed it with the equal weighting of education and occupation. He assigned a nine-point scale to education and a seven-point scale to occupation and broke the resulting index into four discrete social classes (2006 [1966]:171–177).

Trudgill (1974) created his own index whereby he assigned each participant a numerical score according to the following six attributes: occupation (based on the 1966 General Register Office, Classification of Occupations), income, education, father's occupation, housing, and locality. For each attribute, a six-point scale was developed, and participants were assigned a score ranging between 0 and 5. These scores were then added together, rendering a maximum possible score of 30 and a minimum possible score of 0 (1974:35–41).

Critiques of the class index

Rickford (1986:216) cautions that the rigorous ethnographic cross-validation that Warner et al. (1949) subjected their index to has often been ignored by sociolinguists (1986:216). Indices built by Warner et al. (1949) and Hollingshead and Redlich (1958) were not originally meant to be universal. They were highly specific to their respective communities because they were validated according to that community's stylistic expression. In other words, the factor weights developed for New Haven were developed in a regression analysis whose response variable was the subjective impressions of the interviewers (Hollingshead & Redlich 1958:387–397). This is a highly emic validation approach. Further, the cross-validation was done according to media consumption – say, whether a participant preferred the *New York News* (lowbrow) or the *New York Times* (highbrow) (Hollingshead & Redlich 1958:398–407). These are also highly emic artifacts.

Another critique was that the construction of a numerical index begs the question of a stepwise hierarchy. In other words, the conflict between social classes may be more material to the variation of language than gradient access to various forms of capital. Rickford's (1986) research on the Guyanese plantation *Cane Walk* revealed that the conceptualization of social class as a cline was not applicable. The distribution of variables in his dataset was strikingly bifactorial between the subordinated and superordinated groups, which contradicts Bourdieu's (1980) assumption of trajectory. In other words, on a Guyanese plantation, there is really no such thing as peripheral membership to the subordinated group (estate class) in the sense of an 'upper lower class' or 'petite bourgeoisie':

> For the estate class members, however, whose efforts to move upwards within the sugar estate hierarchy (and even outside of it) have rarely been successful, the social order is seen as too rigidly organized in favor of the haves for individual adjustments in language use by the have nots to make much difference. These fieldworkers essentially share a Marxist view of society–some of them explicitly identify it as such–seeing the assigned value of English as just another aspect of ruling class ideology [...]. (Rickford 1986:218)

Although Rickford (1986) doesn't explicitly say it, he is aligned here with Erikson and Goldthorpe (1992) who constructed an index that operationalized class in a way that prioritized relational category over gradient hierarchy, which is much closer to Weberian and Marxist conceptualizations of class. Advocates of this approach to modeling class are critical of scalar SEI and prestige indices and prefer a "categorical versus continuous approach to occupational stratification" (Ganzeboom, De Graaf, & Treiman 1992:3–7).

Macaulay (1976) was as critical of social-class indices as (Rickford 1986), but he did not have a problem with gradient measures, per se. Rather, he cautioned that

> it should not be taken as self-evident that the use of several indicators necessarily provides a more accurate measure of social stratification than the use of a single

> one. This is something which must be carefully justified in terms of the local situation and not something which can be taken for granted on the basis of precedents in totally unrelated situations (Macaulay 1976:185).

This was the approach that Fontanella de Weinberg (1974) took in her study of *-s/s-* aspiration and elision in Bahia Blanca, Argentina. Rather than making an aggregate index *a priori*, she first modeled the variables against occupation and education separately. Six clusters emerged that motivated her creation of a single index (1974:73). Similarly, Labov (2001) also strengthened his analyses by examining how the components of his SES index held up to the linguistic variation he was examining in Philadelphia (2001:113–120,179–186).

Eckert (2012) argues that the use of etic criteria offers an opportunity for replication and comparability despite the obvious shortcoming of ignoring important emic data (2012:90). But even if one is to take the more etic approach, as my study does, one must remember that sociolinguistic samples are notoriously small, which means that even the most 'perfect' index may not result in the 'desired' p-values. For example, this study attempts to make conclusions about Stockholm's speech on the analysis of a mere 36 speakers. If any population-applicable conclusion is to be trusted, then a cumulative analytic approach should be taken. As Fontanella de Weinberg's (1974) does, multiple models should be run on both aggregated and single objective indices. If, after these are run, consistent trends emerge, then the resulting conclusion is more trustworthy.

—

It is not until Chapter 6 that the notions of racialization and class are brought up again. There, I describe and motivate how each is operationalized for Chapter 6 and the related analyses in Chapters 7, 8, and 9. Before, however, beginning the analysis, the proceeding Chaper 5 offers a review of how rhythm has been approached in the literature and how it will be approached in this dissertation.

Citations in original language

[i]"Det är besvärligt eller rent oacceptabelt att tänka sig att dessa rasföreställningar kan förekomma mer än som rena undantag i det svenska folkhemmet, ett modernitetsprojekt som i övrigt präglats av jämlikhet och humanism." (Molina & De los Reyes 2002:297)

[ii]I USA finns idag ett lagförslag i kongressen om att lägga en femte kategori till de fyra existerande raskategorierna: 'multiracial'. Detta visar det hopplösa och absurda i användningen av ett arkaiskt begrepp. (Werbart 2000:183)

[iii][S]å skiljer den [svenska vitheten] ut sig på en avgörande punkt. Det handlar om den exklusivitet som förknippas med den svenska vitheten, och som gör den till den tveklöst mest privilegierade vitheten jämfört med de typer av vithet som återfinns i andra västerländska nationalstater. Det specifika med den svenska vitheten har inte bara gjort att det med stor sannolikhet är svårare att räknas som vit i Sverige än i många andra västländer, utan har även resulterat i att en inhemsk hierarki av olika grader av vithet har växt fram. Den svenska vithetens exklusivitet tar sig bland annat uttryck i att majoritetssvenskar som är kortväxta eller har mörkt hår, mörk ögonfärg och mörkare hud också kan råka ut för rasifiering, misstas för att vara 'invandrare' och få frågor om var de kommer ifrån och varför de ser 'osvenska' ut. [...] [B]åde majoritets- och minoritetssvenskar blonderar håret, använder blåa kontaktlinser,

äter tillväxthormoner eller låter operera sig för att uppnå den idealiserade svenska kroppsnormen. (Hübinette et al. 2012:60)

[iv]Den typiske tyrker adskiller sig fra etniske dansker ved at have en anelse mørkere glød—både når det angår hud og hårfarve. I Tyrkiet er det meget normalt at have en stærk hårvækst. [. . .] Irakere er som udgangspunkt kaukasiske i deres hudfarve. [. . .] Bosnier har en meget lys hud og ligner derfor etniske dansker til forveksling. Deres øjnefarve er ligeledes i den lyse ende af skalaen. Og faktisk har mange bosnier blå eller grønne øjne. (Danmarks Radio 3 2006)

Chapter 5

APPROACHES TO RHYTHM

The study of rhythm is important for understanding language in Stockholm – particularly in the context of its multiethnolect – for two reasons that I will categorize as (1) local dialect description and (2) typology of contact varieties. Locally, a great number of observations have circulated since the 1980s that describe Stockholm's multiethnolect as 'jerky' and 'staccato'. Beyond the case of Sweden, a large body of work on contact varieties has shown that their rhythmic properties resemble a negotiation between superstrate and substrate. The latter topic, along with crosslinguistic typological studies, has been the main catalyst for the literature on analytical approaches to rhythm. In this Chapter I will first address the local literature on rhythm in Stockholm and then move to the literature on contact varieties more generally. I will then review the literature on rhythmic analysis and approaches to rhythm. I close the chapter by identifying how I define rhythm and what analytical approach this dissertation will take.

1 Local dialect description

1.1 *Staccato* Swedish in the immigrant suburbs

The literature on Swedish multiethnolect is sparse compared to the literature on multiethnolects in other languages. Despite their modest number, nearly all of these works have remarked that the variety sounds 'jerky' or 'staccato'. This is all the more notable because the remarks come from non-phonetic works. Kotsinas (1988a, 1990, 1994b), in her groundbreaking work in Rinkeby, was the first to observe that prosody was a salient feature of the variety.

> Even in pronunciation a great deal of features can be described as simplifying, for example a tendency among many speakers to reduce the difference between long and short syllables. Combined with certain prosodic phenomena and sometimes also a clearer-than-normal demarkation between word boundaries makes it such that native Swedes occasionally experience the speech as 'jerky'. (Kotsinas 1988a:268, my translation)[i]

In her later work, she characterizes prosody as the variety's most noticeable feature.

> The most striking feature among many of the speakers, especially A3, D5 and F4, is something described by listeners as 'jerkiness'. What exactly is causing this jerkiness is difficult to determine without a thorough phonetic analysis. (Kotsinas 1990:257, my translation)[ii]

In their ethnography of 'visible boys' and 'invisible girls' in a working-class school in multiethnic Hjällbo (peripheral Gothenburg), Nordenstam and Wallin (2002) identify this feature as the variety's most obvious linguistic identifier as well.

> The most important features that distinguish Hjällbo language is prosody. The way one speaks can be characterized as 'jerky'. This comprises of a stress distribution on both the word and sentence level that differs from the standard language, pitch accents are redistributed or eliminated, and vowels shorten. In aggregate, these features can give a foreign and sometimes monotonous impression. (Nordenstam & Wallin 2002:145, my translation)[iii]

No evidence is offered for their phonetic explanation – neither in the form of citations or the work's analysis – which leads me to conclude it is mostly speculative. It hints, nonetheless, to both the continued salience of prosody in the early 2000s as well as a potential spread of this feature from Stockholm to Gothenburg (or the contemporaneous independent development of this feature in Gothenburg).

Nordenstam and Wallin's (2002) ethnography was released shortly before results began to be published from the first large-scale study on speech in the multilingual neighborhoods

Multilingual suburban secondary-school students Median age: 19	Monolingual university students Median age: 28
'Standard Swedish'	'learner language'
'Swedish	'immigrant' 'Swedish'
'Stockholm dialect'	'refined immigrant Swedish'
'good Swedish'	'suburban Swedish with an accent'
'tries to be smart'	'Rinkeby Swedish'
'good but with a Rinkeby accent'	'suburban prosody'
'better Swedish but with an accent sometimes'	'a staccato-like rhythm'

Table 5.1: An adapted and abridged table of evaluations reported for Ashur's way of speaking from Bijvoet and Fraurud (2011:17)

of Malmö, Gothenburg, and Stockholm, namely the *SUF Project* (see p. 58 for a more thorough review). In one of the earliest outputs from the study, Fraurud (2003) observed also that, alongside lexicon, prosody appeared to be the most defining feature of the variety (2003:87). Although Fraurud's (2003) claim was merely impressionistic, the first-ever acoustic analysis of the prosody of Swedish multiethnolect followed shortly after (Bodén 2004). It demonstrated that both pitch accents 1 and 2 in Malmö's multiethnolect were identical to the received Malmö pronunciation.

As the SUF project came to maturation, it became clear that the ascriptions of 'jerky' and 'staccato' were not just academic observations. Bijvoet and Fraurud (2008, 2011) found that laypeople were making similar metapragmatic evaluations. In their study, two listener groups assessed seven speech stimuli, of which three contained multiethnic youth language. One listener group consisted of monolingual university students. The other listener group consisted of secondary-school students from a working-class multiethnic high school. The authors noted that in particular, one of these three multiethnic-youth-language stimuli rendered quite divergent assessments from the two listener groups.

> This speech sample was recorded in a relatively formal situation (a presentation in front of the class). It contains neither slang words nor grammatical deviations and the pronunciation cannot, according to a panel of linguists from Stockholm University, be traced to any particular first language. It is only on the prosodic level that this speech sample diverges from the dominating regional norm – with its light touch of the 'staccato intonation' often mentioned in descriptions of multiethnic youth language. (Bijvoet & Fraurud 2011:16)

I provide a summary of these diverging evaluations in Table 5.1. It shows only a modest degree of sensitivity among listeners from Stockholm's periphery and a strong degree of sensitivity among listeners from other parts of Stockholm.

Not only do the right-side evaluations in Table 5.1 imply some level of sensitivity to staccato rhythm, the results also imply non-uniform levels of sensitivity. This in turn could imply that staccato may be in the intermediate stage of evolution, referred to by Labov (2001:196) as a linguistic *marker*.

> As they proceed to completion, such changes usually acquire social recognition as linguistic *markers*, usually in the form of social stigma, which is reflected in sharp social stratification of speech production, a steep slope of style shifting, and negative responses on subjective reaction tests. (Labov 2001)

The topic of salience, markers, and contextual style will be discussed in detail in Chapter 8 (p. 207).

There are no indications in Bijvoet and Fraurud (2011) that rhythm is a feature that speakers style-shift out of, but the responses in Table 5.1 indicate perhaps some social stigma, albeit no overt negativity. In a later examination of their data, Bijvoet and Fraurud (2016) found that the presence of 'staccato' rhythm for speaker 'Eleni' was sufficient to make listeners think they had heard grammatical errors that were not actually there. They also note that even though Eleni sheds a number of 'suburban' features when she style-shifts, she does not quite have access to her prosody.

> It is on the prosodic level that her way of speaking with its light and shifting touch of the 'staccato intonation' that is often associated with suburban young people (Bodén 2011; Kotsinas 1994b, 2000) slightly diverges from the dominating monolingual regional norm. However, a majority of listeners in a pilot study labelled a sample of Eleni's speech 'Rinkeby Swedish', and several of them also counterfactually found her speech to be ungrammatical and to contain slang words (Bijvoet & Fraurud 2011). Evidently, a single prosodic feature associated with a low-status variety was enough to make listeners also hear what was in reality not there.

The notion that just surface prosody could elicit strong listener responses was later supported in my earlier work (Young 2014, 2018c). Two listener groups evaluated eight audio recordings of working-class men making a table reservation at a restaurant. One listener group assessed the stimuli as either belonging to a 'multiethnic neighborhood' or a 'Swedish neighborhood' while the other listener group assessed them on a seven-point Likert scale with 'very rough' on one end, 'neutral' in the middle, and 'very refined' on the other end. Stimuli assessed as 'multiethnic neighborhood' and 'rough' had lower intervocalic alternation (i.e., 'staccato'). Stimuli assessed as from a 'Swedish neighborhood' and 'neutral' or 'refined' had higher intervocalic alternation (Young 2018c:50). These results, like Bijvoet and Fraurud's (2011), also implicated 'staccato' as a feature garnering negative subjective reactions. Furthermore, there were indications in this study that some speakers can attenuate the staccato feature when they style-shift. Two speakers assessed positively and as standard by listeners (pseudonyms 'Reman' and 'Samir')[1] did not typically use standard Swedish in casual speech

[1]Note that speaker Reman is also a participant in this dissertation.

(Young 2014:9,19). This, however, was my subjective assessment; I did not actually analyze casual speech, and no evidence was provided for this claim.

Prosody and style-shifting emerge as a more explicit topic in Milani and Jonsson's (2012) ethnographic study of youth in a multiethnic suburb of Stockholm. Participant 'Emre' relays an encounter with a police officer whereby he shifts out of his vernacular register into the normative Stockolmian style, shedding 'staccato' rhythm.

> On the one hand, the performance of 'the policeman' is rendered with a lower pitch and an easily recognizable southern inner city Stockholm accent. On the other hand, his own answers are recounted with the 'staccato-like' rhythm associated with Rinkeby Swedish. (Milani & Jonsson 2012:54)

—

The above accounts complete the literature in and on Sweden in regards to rhythm in multiethnolect. They imply that (1) some sort of variable is operating to give the impression of 'jerkiness' or 'staccato' and (2) that there is some degree of social salience. Testing these two implications constitutes the main purpose of this dissertation. The following section will move away from local observations and review the literature on rhythm in contact varieties more generally.

1.2 Rhythm in Stockholm/ Central Swedish

Much of the research on rhythm in Swedish has focused on Central Swedish, which is the regional variety to which Stockholm Swedish belongs (Riad 2014:8). Some of the findings in this section rely on terms that are first covered in the review of rhythm later on in Section 2 (pp. 105–119).

Strangert (1985) conducted the first crosslinguistic comparison of Swedish rhythm by comparing it to Spanish and Finnish. The study involved five speakers of Stockholm Swedish, one speaker of Finnish, and one speaker of Spanish. Its main goal was to test the application of the *isochrony principle* (see explanation in next section) to the three languages rather than to examine rhythmic alternation, per se. Nonetheless, some of her findings imply that Swedish might have higher alternation than Spanish. For example, "the ratios of monosyllabic intervals to intervals with two and three syllables in Swedish amounts to $1 : 1.17$ and $1 : 1.57$, respectively, and the corresponding ratios for Spanish are $1 : 1.30$ and $1 : 1.80$" (Strangert 1985:133). The results imply that Swedish sits with other generally high-alternation languages like Dutch and English (Dauer 1983; Grabe & Low 2002; White & Mattys 2007).

Lindblom and Rapp (1971) and Lindblom (1975) conducted a much smaller study that was neither cross-linguistic or cross-varietal. They tested Central Swedish-speaking participants with a series of fabricated words that increased in syllable-count to test whether words increased in duration in linear proportion to syllable count. The authors found they did.

While the above studies were mainly focused on foot structure, Bruce (1983) proposed that Swedish rhythm might operate epiphenomenally, beyond even the dynamic of prominence. In his pilot study he found that the unstressed syllables between two stressed syllables in a prosodic phrase fell into an alternating pattern of *weak-strong-weak*. In situations where the number of unstressed syllables between two stresses was even, the pattern prioritized maintaining the final strong-weak pattern. A single participant was asked to read aloud a sentence and then to immediately repeat the intonational pattern of the sentence with the syllables *ba* and *da*. The participant read each of the below sentences between five and seven times, the recordings were time-aligned, and the durations were measured.

ˈkompisar med ˈtjeckerna	*ˈba da ba da ˈba da ba*	$+ - - - + - - \rightarrow$	$+ - \underset{+}{-} - + - \underset{+}{-}$
ˈkompisar med solˈdater	*ˈba da ba da ba ˈda ba*	$+ - - - - + - \rightarrow$	$+ \underset{+}{-} - \underset{+}{-} - + -$
ˈkompisar med kapiˈtal	*ˈba da ba da ba da ˈba*	$+ - - - - - + \rightarrow$	$+ - \underset{+}{-} - \underset{+}{-} - +$
solˈdater med kapiˈtal	*ba ˈda ba da ba da ˈba*	$- + - - - - + \rightarrow$	$- + \underset{+}{-} - \underset{+}{-} - +$
attenˈtat med kapiˈtal	*ba da ˈba da ba da ˈba*	$- - + - - - + \rightarrow$	$\underset{+}{-} - + - \underset{+}{-} - +$
attenˈtat med solˈdater	*ba da ˈba da ba ˈda ba*	$- - + - - + - \rightarrow$	$\underset{+}{-} - + \underset{+}{-} - + -$
attenˈtat med ˈtjeckerna	*ba da ˈba da ˈba da ba*	$- - + - + - - \rightarrow$	$\underset{+}{-} - + - + - \underset{+}{-}$

As shown above, Bruce found that the speaker typically lengthened every second unstressed vowel (marked with $\underset{+}{-}$), rendering a weak-strong-weak alternation. For even numbers of unstressed syllables, the alternation pattern prioritized beginning with strong and ending with weak. Despite its small sample size, the study is significant because of its gutsy theoretical proposal. Bruce proposed that rhythm may operate independently of notions of the prosodic foot and independently of its surface-segmental constitution:

> It is my bias in thinking about rhythm in spoken language that the leading principle is alternation and not isochrony. [...] The principle of rhythmic alternation entails a conspiracy against sequences of several equally prominent syllables. This applies to both unstressed and stressed syllables. (Bruce 1983:36–37)

As far as I am aware, Bruce did not expand the study to more participants; nor was the study ever duplicated by other researchers.

Fant and Kruckenberg (1989) took a more traditional approach to Central Swedish speech rhythm and investigated whether read-aloud speech was isochronous; in other words, whether the duration of a prosodic foot remains constant regardless of how many syllables are within it. They found this not to be the case (1989:57). Nonetheless, they offered that there might be a rhythmic component to the constitution of pauses between breath groups. This is because they found that in rhythmic reading, the sum of the phrase-final syllable and the subsequent pause equalled the mean duration of the prosodic foot for each speaker (1989:58).

Eriksson (1991) conducted a study similar to Lindblom and Rapp (1971) and Lindblom (1975), albeit with a much larger sample size. His main focus was also to test the isochrony principle in Stockholm Swedish (whether words increase in duration proportional to syllable count; see p. 105). He time-aligned and extracted the syllable durations from five sentences read aloud by 30 participants and found that the duration of prosodic feet did increase in linear proportion to the addition of syllables. This finding was in agreement with both Lindblom and Rapp (1971) and Lindblom (1975) as well as Fant and Kruckenberg's (1989) resolute conclusion that Swedish is not isochronous (1989:57).

Krull (2012) conducted the first cross-varietal study of rhythm in Swedish. She found that female speakers of Central Swedish ($n = 5$) had much higher rhythmic alternation than L2 Swedish spoken by L1 Estonians ($n = 5$). Asu, Nolan, and Schötz (2017) analyzed recordings of read-aloud speech from three speakers of the Nücko variety of Estonian Swedish from 1920. They compared them with three elderly Central-Swedish speakers but found no significant inter-varietal difference for rhythmic alternation (2017:76). Further, although Krull (2012) and Asu et al. (2017) used the same metrics in their analyses (nPVIV; Low et al. 2000, reviewed on p. 115), their results for Central Swedish were quite different.

Motivated by the ascriptions described in the aforementioned section, I conducted a small perception study of the speech of men in Stockholm (Young 2018c). I found that the four speech samples assessed as multiethnolect had significantly lower alternation than the four speech samples assessed as standard. I used the same metrics as Krull (2012) and Asu et al. (2017), but only Krull's (2012) measurements for Central Swedish were consistent with my findings for standard Stockholm Swedish (2018c:50). When the metrics from Krull (2012) and Young (2018c) for Central/Stockholm Swedish are compared with investigations of other languages, Swedish falls as one of the highest-alternation languages in the world, below British English (Fuchs 2013) and just slightly higher than American English (Thomas & Carter 2006).

1.3 Relevant areas outside of Stockholm

Similar observations of the supposed 'jerky' attribute have been made regarding the multiethnolects in Gothenburg and Malmö (Sweden's second and third cities, respectively). In an attempt to identify the source of this impression, Bodén (2007, 2010) compared vocalic duration between the received Malmö standard and the local multiethnolect (aka *Rosengård Swedish*) in a small random sample of her speaker corpus. No significant difference was found in the sample. In other words, phonologically-long vowels did not shorten, and phonologically-short vowels did not lengthen with respect to the standard Malmö Swedish samples (2007:29). She hypothesizes that the absence of elisions and reductions might be responsible for this (2007:33).

Unlike her quantitative investigations on rhythm, Bodén (2007) successfully identifies some qualitative differences in intonation between standard Malmö and Rosengård Swedish. In the phrase 'ida(g) vi lovade' (today we promised), typical Scanian intonation would gradually decline across the phrase and include two peaks and troughs. In the multiethnolectal style, declination was often absent. Additionally, three contours occur in the multiethnolectal phrase instead of two in the standard. This particular phrase was taken from an informant who wanted to demonstrate the intonational difference between Rosengård Swedish and standard Scanian Swedish (Bodén 2007:36-37). Given the salience this intonational pattern held for that informant, the additional contour is something that might be a mature feature in Malmö's multiethnolect.

Bodén (2010) conducted further work on prosody in a qualitative analysis of multiethnolects from all three major cities in Sweden with a particular focus on Malmö and Gothenburg. Importantly, she found that select speakers of multiethnolect in both Malmö and Gothenburg had durationally longer phrase-final syllables than speakers of the respective regional mainstream varieties (2010:72–75). She also examined one multiethnolectal speaker from each of the three cities, including Stockholm, and found that they had an "expanded *fo* range" compared to the standard styles in each of the three cities (2010:75).

Hansen and Pharao (2010) observed that Copenhagen multiethnolect was 'staccato-like' and investigated the source of this rhythmic difference, concluding it was due to transformations of certain durational categories of vowels. Danish, like Swedish, has phonologically long and short vowels (reviewed on pp. 43–45). Twelve speakers of multiethnolect and 12 speakers of standard Copenhagen Danish participated in a map task that targeted 33 test words. The authors found that speakers of multiethnolect generally had "equal duration of long and short vowels before syllables containing a full vowel." (2010:93) and that this was generally "due to shortening of long vowels rather than lengthening of short vowels." (2010:91). Hansen and Pharao's (2010) proposal will be tested on this dataset in Chapter 7.

Fagyal (2010) was motivated by similar observations of 'staccato rhythm' in Verlan (Cerquiglini 2001; Duez & Casanova 2000) when she investigated the speech rhythm of young working-class boys in a Parisian suburb. Although no significant rhythmic differences were found between youth of North African descent and youth of European descent, small segmental differences emerged. Youth of North African descent had slightly lower vocalic and higher consonantal durations than youth of European descent (2010:117). In a later examination of the same dataset, Fagyal and Stewart (2011) found that "that phrase-final rising-falling intonation, typical in certain types of imperative in French, has a much broader pragmatic meaning in working-class youth vernacular where it seems to function as a micro-level style feature indexing common ground and in-group affiliation with members of the adolescent peer group" (2011:75). This finding connects with the analysis that I later conduct in Chapter 7 on phrase-final lengthening (p. 188, p. 202).

In London, Torgersen and Szakay (2012) provided convincing evidence that London's multiethnolect (Multicultural London English, MLE) in Hackney has less intervocalic durational contrast than the white working-class varieties in Havering. Just as Hansen and Pharao (2010), they propose that the difference is phonotactic and originates from specific segmental transformations.

What may be part of the cause of changes in the rhythmic patterns is a change

> in the duration of particular segments. We have shown how FACE, and to some degree GOAT, have shorter durations in Hackney than in Havering. The non-Anglos are in the lead in having near-monophthongal variants and also have a more syllable-timed rhythm than the other speakers. (Torgersen & Szakay 2012:838)

They propose that the lower alternation in MLE is connected to multiethnolectal speakers using monophthongal vowels, a notion I explore in Chapter 7 (pp. 189–200).

2 Typological approaches to rhythm

2.1 The isochrony principle

Much of the contemporary literature on rhythm in language has defaulted to the term *timing*, often conflating the latter with the former (Arvaniti 2009). I follow Arvaniti (2009) in treating them as separate constructs with inevitable ties to each other. This means that an exhaustive review of the literature on rhythm will include a review on timing and, to begin with, the isochrony principle.

The literature on linguistic timing is as old as the notion of the syllabic foot itself. The *isochrony principle* emerged from a Rationalist-era epistemology that sought to create a rigid symmetrical taxonomy for most natural phenomena. This meant accounting for speech prosody via the logic of verse (see Adams 1979:22–28 for an excellent review). The underlying and mostly unchallenged claim within this account was that speech, specifically English [2], conformed to a symmetrical grid in which the duration between accents or stresses was always equal, regardless of the number of syllables (or pauses) between them. For example, *like* would have the same duration as *unlikely* because the syllables in *unlikely* would undergo compression. The earliest record of the isochrony principle is typically attributed to Steele (1775 [1969]):

> [W]e must presuppose an exact periodical pulsation, as regular as the swings of a pendulum, the length of which periodical pulsation we may vary according to our pleasure, as often as we would chuse to quicken or slacken the movement; and then all continuation of sounds or pauses are to be subserviently measured and regulated by this uniform and steady pulsation, as long as that proportion of pulsation (or pendulum) shall be continued (Steele 1775 [1969]:67).

Steele asserts here that sound components are subordinate to and adhere strictly to the exact periodicity of measure. The claim reflects a "Kantian emphasis on universalizable principles" (Donaldson 1992:137) and insinuates a Classical association between morality and order.

[2] As far as I am aware, Brücke (1871) was the first scholar to discuss isochrony in the context of a non-English language.

> For the Greeks, who regarded rhythm as the application of order to movement, a correspondence existed between ordered movement and the movements of the soul – thus rhythm was seen to have an ethical function. [...] Thus, as well as denoting harmony or symmetry in its more general sense of the right relation of the parts to each other and the whole, rhythm for them was fundamental to their theory of the good and the beautiful, and as such a distinctly human characteristic. (Adams 1979:9)

The two most-cited contemporary studies on rhythm attribute the isochrony principle to Pike (1945) (Low et al. 2000; Ramus et al. 1999), but the reasoning behind the principle is older and more established than that. Adams (1979) points out that Odell (1969 [1806]), Chapman (1972 [1818]), Patmore (1857), and, importantly, Thomson (1904) have all promoted the notion of isochrony.

> Within certain limits, therefore, the insertion or omission of unaccented syllables does not affect the total duration of a phrase, and the length of accented syllables varies according to the character and number of unaccented syllables intervening before the next accent. (Thomson 1904:65)

In the aforementioned excerpt, Thomson (1904) provides yet another unique explanation to an already well-established idea. Such regenerations of the isochrony principle continued into the late 1960s, which may explain why it has become so difficult for scholars to accept the more recent research that has disproven it (Dauer 1983; Duckworth 1967; Lehiste 1972; Lindblom 1975; O'Connor 1965; Shen & Peterson 1962; Strangert 1985).

Three mid-century works are often cited by modern-day authors as the originators of the isochrony principle: Classe's (1939), Pike's (1945), and Abercrombie's (1967). Famous for his contributions in other subfields, Pike (1945) is by far the most famous of the three. However, Pike's claims were actually based on an oversimplification of Classe's (1939) findings. Classe, himself, acknowledges that his dataset does not fully support the isochrony principle:

> From the very nature of speech, it is obvious that, in the normal course of events, all the necessary conditions will generally not be present at the same time. But it does not follow that isochronism cannot be the *basis* of the rhythm of English. If we consider some of the less isochronous groups, we shall notice an interesting fact. Suppose in the first place that the duration of a group depends exclusively on the nature of the component syllables, their number, and the grammatical context of the group. If no other influence made itself felt we should then expect the duration of the group to be more or less proportional to the number of syllables and the importance of the 'cuts'. But this is not the case[...]. Whenever short groups are mixed with longer ones, the speaker minimizes the differences by changes in his rate of delivery. [...] It is not inconsistent with the facts to ascribe the acceleration, when it is present, to a desire on the part of the

> subject to equalize the groups, or to speak of a *tendency towards* isochronism (Classe 1939:86–87).

In the above quote, Classe argues that there is to a weighted hierarchy by which isochrony loses out to surface features. This *tendency* towards isochronism presumes that surface components like segmental inventory, semantic prominence, and lexical stress are complications to an epiphenomenal isochronic base (89), which is a fascinating and thought-provoking argument. Although the ensuing studies disprove isochrony as any notion, Classe's (1939) idea of an epiphenomenal rhythmic grid is supported in part by my findings in Chapter 7.

As I prefaced above, the isochrony principle is nearly always attributed to Pike (1945). According to Adams (1979:49), this is primarily because he introduced the contrastive 'upset'; namely, the notion of syllable-timed rhythm.

> Since the rhythm units have different numbers of syllables, but a similar time value, the syllables of the longer ones are crushed together, and pronounced very rapidly, in order to get them pronounced at all, within what time limitation.[...] Because its length is largely dependent upon the presence of one strong stress, rather than upon the specific number of its syllables, it may conveniently be labelled a STRESS-TIMED rhythm unit. Many non-English languages (Spanish, for instance) [...] tend to come at more-or-less evenly recurrent intervals–so that, as a result, phrases with extra syllables take proportionately more time, and syllables or vowels are less likely to be shortened or modified [...], the unit becomes somewhat STACCATO. (Pike 1945:35, original emphasis)

Like in the descriptions of Swedish multiethnolect in Section 1.1, Pike also uses the term 'staccato' to describe Latin-American Spanish, his main subject of study. In my view, it is hard to avoid critically considering the iconicity that the stereotypes of the measured English and unmeasured Latin-American share with *stress-timing* and *syllable-timing*. In fact, the substantiation that Pike offers for stress and syllable timing is the following statement.

> The recurrent stress timing is perhaps even more important than the number of syllables in iambic or trochaic groups, or the like. Evidence of this fact is seen in the esthetic satisfaction obtained by English speakers from some lines of poetry [...]. [34](Pike 1945)

Nonetheless, the proposal was appealing and gained tremendous traction. For decades, many scholars, including sociolinguists, regarded it as self-evident and ignored the many studies that disproved his hypothesis. Dauer (1983) observed that

> The concept of syllable-timing was originally developed by English speakers to describe a rhythm that is *opposite* to that of English, that is, it has been defined primarily negatively. However the label has not been widely accepted by native

speakers of those languages described as such. Navarro (1932:206–207) strongly denies that syllables are of equal length in Spanish, and asserts that Spanish verse is based on the regularity in length of intervals between rhythmic accents. Olsen (1972:996) cautiously calls Spanish syllable-timed "even though it has covering patterns of rhythm based on sequences of stressed and unstressed syllables". Grammont (1932:105) considers rhythm in French to be based on the recurrence of accented syllables [...]. (Dauer 1983:60, original emphasis)

Eriksson (1991) noted that

> [Pike's] ideas are put forward in the form of statements, much like those of traditional style normative grammars. They are not presented as hypotheses meant to be subject to empirical tests, or the result of empirical studies, by himself or by others. Reading his book today one finds it surprising that his ideas got the attention they did, but for whatever reason, they *did*, and they influenced speech rhythm research to a considerable extent in the following decades. (Eriksson 1991:19, original emphasis)

Before I move to the evidence against the isochrony principle, I will review one last key proponent of isochrony who, like Pike, is often cited as the originator of the terms *syllable* and *stress timing*. Abercrombie (1967) claimed that the rhythmic impression of a language results from the timing difference in small puffs of air called *chest pulses* and that "the respiratory muscles alternatively contract and relax at a rate of roughly five times per second, so that the air is expelled in a succession of small puffs. Each contraction, together with the resulting puff of air, constitutes the basis of a syllable." (1967:35). When a chest pulse, however, "is produced by exceptionally great muscular action, [...] the extra strong muscular movement itself is called a *stress-pulse*." (Abercrombie 1967:35) These definitions, while unsubstantiated, formed the basis for a biological embellishment of isochrony theory:

> When one of the two series of pulses is in isochronous succession, the other will not be. Thus in a syllable-timed rhythm, the stress-pulses are unevenly spaced, and in a stress-timed rhythm the chest-pulses are unevenly spaced. (Abercrombie 1967:97)

2.2 Testing the isochrony principle

Although the isochrony principle was disproven in a series of late twentieth-century studies, much contemporary laboratory-phonology and sociophonetic work relies on Pike and Abercrombie for their theoretical premise. According to Adams (1979:49), "Pike's authoritative statements did much to propagate the belief that isochrony constitutes the basis of English speech rhythm", and according to Arvaniti (2009:46), "the stress-/syllable-timing distinction received renewed interest with the advent of rhythm metrics".

Words		*'Stressese'*				Total foot duration (ms)	*'Syllabese'*				Total foot duration (ms)
1 syllable	like	ˈlike					ˈlike				
Syllabic durations (ms)		150				150	90				90
2 syllables	likely	ˈlike	ly				ˈlike	ly			
Syllabic durations (ms)		120	30			150	90	30			120
3 syllables	likelier	ˈlike	li	er			ˈlike	li	er		
Syllabic durations (ms)		90	30	30		150	90	30	30		150
4 syllables	unlikelier	un	ˈlike	li	er		un	ˈlike	li	er	
Syllabic durations (ms)		30	60	30	30	150	30	90	30	30	180

Table 5.2: Hypothetical languages based on Pike's (1945) dichotomy of timing in language. 'Stressese' synthesizes a perfectly stress-timed language, and 'Syllabese' synthesizes a perfectly syllable-timed language.

Therefore, it is crucial to understand exactly what its claim is, what has been refuted, and what hasn't been refuted.

To do this, I first provide Table 5.2 with two hypothetical languages that I coined, *Stressese* and *Syllabese*, which fit Pike (1945) and Abercrombie's (1967) paradigm. In Stressese, the agglutination of *like* shortens the root morpheme in order to ensure that the duration of the foot remains unchanged[3]. Syllabese, by contrast, incrementally lengthens the foot in linear proportion to every unstressed syllable. Hypothetical durations are provided to show the potential implications each have on rhythm: In Stressese, the total duration remains at 150, and each additional syllable reduces the durational contrast between the stressed and the unstressed syllable. In Syllabese, the duration changes in linear proportion to the number of syllables, from 90 to 180 for one to four syllables, respectively. The durational contrast here remains stable.

Subsequent work failed to demonstrate empirically that isochrony existed in English (or any other language). According to Adams (1979:49–51), three studies also coincided with and refuted Abercrombie's (1967) reassertion of isochrony[4]. Shen and Peterson (1962) conducted the first investigation of isochrony that made use of spectrographic analysis. "We did not find isochronism in our limited data and therefore cannot say that there is isochronism in English" (1962:34). Not only did O'Connor (1965) find that "there was no physical isochrony between the four groups in each line" and that "the phonetic structure of the word influenced the duration of the stress groups", he found that limericks did not adhere to the isochronic principle either (1965:11). Duckworth (1967) drew a similar conclusion:

> In the sample studied, the only conclusion can be that the notion of stress timing or isochronism which has persisted for a number of years, could not be defended. The number of syllables per segment[5] seemed to be a much more

[3] It is unclear whether ideal syllable-timed languages also compress unstressed syllables or not. In this example, I only compress the stressed syllable.

[4] Ladefoged (1967), Lebrun (1966), and Adams (1979) also disproved the *chest-pulse* hypothesis.

[5] In this context, 'segment' refers to the stretch of speech examined.

powerful determining factor in the duration of a segment than the placement of stress. (Duckworth 1967:12)

2.3 From isochrony to compressibility

Three years after Duckworth's (1967) study, O'Connor (1968) attempted to isolate the effects of an isochronic base by having participants put together increasingly long non-sensical sentences: they were to fill "Take ____ Park" with the fictional words *ses*, *sets*, *sekts*, *seksts spektsts*, *spreksts*, and *spreiksts* (O'Connor 1968). Although the results showed an increase in duration as segmental length increased, "the duration [was] not directly proportional to the number of segments; there is therefore a compressive tendency which might correspond to the tendency to isochronism mentioned in phonetic literature." (O'Connor 1968:1) Notably, the findings support Classe's (1939) original postulate that there is "a desire on the part of the subject to equalize the groups, or to speak of a tendency towards isochronism." (86)

Bertinetto (1989) refers to this compressive tendency as *compensatory shortening* (1989:113). Two subsequent studies appeared to confirm this same sort of tendency. Lehiste (1972) found that English stressed syllables decreased in duration when subsequent unstressed morphemes were added (*stick*, *sticky*, *stickier*, *stickiness*), but he did not find a reciprocal reduction of unstressed syllables to maintain a durationally-controlled foot.

Lindblom and Rapp (1971) and Lindblom (1975) tested Swedish participants with a series of made-up words that increased incrementally in syllable-count and found the same thing. A key problem, however, is that these and the above findings call into question the dichotomy because their results match stress-timing in some respects and syllable-timing in others. Although the duration of stressed vowels do decrease with syllable-count – 278, 273, 270, 255 millisecond for one, two, three, and four syllables, respectively – they only do this mildly (1971:21). One could therefore make the argument that they are closer to the 90-90-90-90 model for unstressed syllables in Table 5.2's *Syllabese* than the 150-120-90-60 model in *Stressese*. Unstressed syllables barely compress at all, going from 160 to 155 to 150 for 2, 3, and 4 syllables respectively (Lindblom 1975:394), again implying that the relationship between foot duration and syllable count in Swedish is closer to Syllabese than to Stressese.

Balasubramanian (1980) also found that unstressed syllables barely compress in Tamil. Crompton (1980) found a similar lack of compression in French (1980:223). Manrique and Signorini (1983) did not find much compression in Argentinian Spanish either and, importantly, conclude that it is not very different from English. Nakatani, O'Connor, and Aston (1981) found "no evidence to support even a liberal interpretation of isochrony." (1981:103). Their analysis was of read-aloud speech rather than single words, and there was a near-perfect linear association between syllable-count and durational increase (1981:89).

O'Connor (1968), Lehiste (1972), and Lindblom and Rapp's (1971) findings on foot structure look like a compromise between *Syllabese* and *Stressese*. Therefore, the question arises whether English is closer to *Stressese* than other languages like, for example, Spanish.

This was the same question Dauer (1983) had when she conducted the first systematic comparison of multiple languages in an attempt to identify isochronic properties and/or differences. She recorded readings in English ($n = 4$), Greek ($n = 3$), Italian ($n = 2$), Spanish ($n = 3$), and Thai ($n = 1$) and extracted durational data for their syllables. No language, English included, showed consistency in foot duration. Further, she found no significant differences in mean durations or standard deviations between the languages compared. Importantly, durations increased as a linear function of the number of syllables for all of the languages in the study, and compression was found for all of the languages. Given the preceding discourse that focused on English and Spanish as emblematic stress and syllable-timed languages, respectively, the most important finding here was that their foot structures were not significantly different.

Roach (1982) conducted a nearly identical study whereby Arabic, English, French, Telugu, Russian, and Yoruba were examined, yet with only one speaker each. Like Dauer's (1983) results, interstress intervals were not significantly different. Strangert (1985) similarly conducted a comparative study of Swedish with Finnish and Spanish and found that "in spite of certain differences between the speakers of the three languages, they all exhibited similar temporal adjustments" (Strangert 1985:137). Vayra, Avesani, and Fowler (1984) duplicated Lindblom and Rapp's (1971) study for Italian and compared it to Lindblom and Rapp's (1971) Swedish results. They found compensatory shortening to be much lower, but their study was a pilot with just one speaker. Farnetani and Kori (1983) found similar results with more speakers.

The fact that all languages showed compensatory shortening, yet in various degrees, led to a reconceptualization of isochrony altogether. Bertinetto (1989) suggested the term "compressibility":

> It follows that the original dichotomy has gradually lost much of its dichotomic character, and has more and more acquired the aspect of a scalar orientation. [...] Actually, one might even entertain the idea that the very opposition 'stress- vs. syllable-timing' is somewhat misleading: a single term (such as 'compressibility'), connecting both ideal types along one and the same dimension, might serve the purpose. (Bertinetto 1989:112)

2.4 From compressibility to rhythm

Dauer (1983) is generally credited with moving the discussion from isochrony and durational compression to the realm of rhythm. As cited before, her analysis of English, Greek, Italian, Spanish, and Thai showed neither isochrony nor a difference in compression among the languages. In the same study, however, she conducted a post-hoc comparison of the phonetic structure in English and Spanish and proposed that three phonetic features were independently operating in each language to give the perception of stress and syllable-timing.

English				Spanish			
Stressed		Unstressed		Stressed		Unstressed	
CVC	35%	CV	38%	CV	53%	CV	61%
CV	28%	CVC	27%	CVC	24%	CVC	22%
CVCC	13%	VC	20%	CCVC	12%	V	7%
CCVC	6%	V	10%	CCV	9%	CCV	5%

Table 5.3: Reproduced from (Dauer 1983:57): Most frequently occurring stressed and unstressed syllables in English and Spanish

First, she found that Spanish was dominated by CV-syllables whereas English had a more even distribution of CV, CVC, VC, and CVCC syllables (Dauer 1983:56). These results are reproduced in Table 5.3. Second, 92% of unstressed CV syllables in English were composed of a consonant plus /i/, /ə/, or /ɚ/ versus whole vowels in stressed CVs. In Spanish, there was nearly no quality difference at all between unstressed and stressed vowels (1983:57). Third, the durational properties of stressed syllables are significantly different between the two languages. In English, stressed syllables are on average 1.5 times longer than unstressed syllables, and in Spanish this ratio is 1.3. She therefore proposes replacing the term 'timing' with *rhythm*.

> If we believe that the rhythmic types characterized as stress-timed and syllable-timed really have nothing to do with timing, in the sense of equality of interstress or intersyllable intervals, are we justified in using these terms at all? Both Allen (1975) and O'Connor (1973) have used the term 'stress-based' languages. We can say that a language is more or less stress-based, depending on how large a role stress plays in that language, especially with respect to the three areas described above. [...] Languages can be compared to each other along the dimension of having a more or less stress-based rhythm:
>
> ——————————————————→ Stress-based
>
> Japanese French Spanish Greek Portuguese English

The key difference here between her proposal and Bertinetto's (1989) is that Bertinetto (1989) failed to acknowledge one key fact. The studies that claimed to show differences in compression were few and thin; in other words, they were based on very small datasets. In contrast, as Dauer (1983) indicates in her above citations, there are clear accounts of stress and prominence varying by language alongside clear differences in *syllable constitution*, demonstrated in Table 5.3.

Eriksson (1991) conducted a similar study for Swedish. He time-aligned and extracted the syllable durations from five sentences read by ten adult men, ten adult women, six boys, and four girls. Like the aforementioned studies, he found that Swedish is not isochronous and the duration of words increase in linear proportion to the addition of syllables. However, unlike the prior studies on Swedish, he found no evidence of compression (1991:202). Note, however, that his conclusions do not differ from these earlier studies. The earlier studies found only *marginal* amounts of compression (Lindblom 1975; Lindblom & Rapp 1971). Importantly also, the small amount of compression was not greater for Swedish than for other languages (Strangert 1985:137).

2.5 The 'metrics era': Crosslinguistic and crossvarietal analyses

Over ten years passed between Dauer's proposal and the next serious study of rhythm. Ramus et al. (1999) broke this hiatus with their seminal investigation of rhythm in Catalan, Dutch, English, French, Italian, Japanese, Polish, and Spanish. Ramus et al. (1999) were also the first among crosslinguistic and crossvarietal studies of rhythm that relied exclusively on mathematical metrics. The crosslinguistic studies are typified by an attempt to place a number of languages along an alternation cline. The crossvarietal studies typically involve a comparative analysis between a contact variety – sometimes a multiethnolect – and its mainstream counterpart.

The metrics can be divided into two categories: *global* and *local* (Fuchs 2013:30). The difference between these two categories will become evident in the following sections.

Global metrics (%V, ΔV, ΔC, VarcoV, VarcoC, DSI)

Ramus et al.'s (1999) study was the first on two fronts:

1. Inspired by Dauer's (1983) findings on syllable constitution, it used mathematical metrics to measure the ratio of vowel segments to total segments and the standard deviation of vocalic and consonantal durations

2. While discrediting isochrony theory in its review, it revived the long-discredited terms 'stress timing' and 'syllable timing'

Four speakers were elicited for each language ($n = 8x4 = 32$) to read five declarative sentences. The recordings were time-aligned and segmented according to vocalic and consonantal intervals: e.g., 'next Tuesday on' became C-V-C-V-C-V-C from /n-ɛ-kstj-u-zd-eiɔ-n/ (Ramus et al. 1999:272). Three metrics were calculated. The sum of vocalic intervals were divided by the total duration of the sentence (%V), and the standard deviation of consonantal durations (ΔC) and standard deviation of vocalic durations (ΔV) were calculated:

$$\Delta = \frac{\sqrt{\sum(d_k - \bar{d})^2}}{n} \tag{5.1}$$

where d_k = duration of segment at k^{th} position
where $\bar{d}$ = mean duration of segments
where n = number of segments

Neither ΔC nor ΔV rendered very coherent results, but %V did, illustrated in Table 5.4. The authors assessed this suitability of %V by means of how close it resembled the *a priori* ranking they had established based on generally-accepted preconceived notions of where the languages fit.

Languages	%V (st. dev.)
English	40.1 (5.4)
Polish	41.0 (3.4)
Dutch	42.3 (4.2)
French	43.6 (4.5)
Spanish	43.8 (4.0)
Italian	45.2 (3.9)
Catalan	45.6 (5.4)
Japanese	53.1 (3.4)

Table 5.4: Reproduced from (Ramus et al. 1999:272): Proportion of vocalic intervals (%V) averaged by language and their respective standard deviations

Ramus et al.'s (1999) metrics were the beginning of a distinct era in rhythm studies. Nearly every subsequent study incorporates these metrics, iterations of these metrics, or iterations of iterations of these metrics. The second way in which Ramus et al. (1999) changed the field is that they revived the terms *stress* and *syllable-timing*. As I mention above, they themselves reject the isochrony principle, but the nomenclature that they revive makes use of the term *timing* instead of *rhythm*, something that inevitably implicates isochrony.

Local metrics (nPVIV, rPVI-V, rPVI-C, RR)

The development of local metrics reflects a refinement in how mathematics is used to model speech phenomena. Rhythm is constructed by means of the relationship between elements that have relative proximity to each other. These metrics attempted to take that into account. According to Fuchs (2013),

> Some of the metrics are computed globally, by taking into account how all the intervals in an utterance or a sentence differ from each other. The standard deviation is an appropriate statistical measure for this task. Because differences in duration are particularly salient when long and short intervals follow each other, local rhythm metrics can capture durational characteristics that elude the global ones. (Fuchs 2013:30)

Fuchs' critique of global metrics is due to their internal properties rather than their success in confirming *a priori* assumptions about sundry languages. The problem with global metrics like standard deviations is that rapid speech renders a smaller standard deviation than slow speech, regardless of actual contrastive variation. Consider the standard deviation of 100-50-100-50-100-50 versus 50-25-50-25-50-25; both are arguably very similar contrasts with

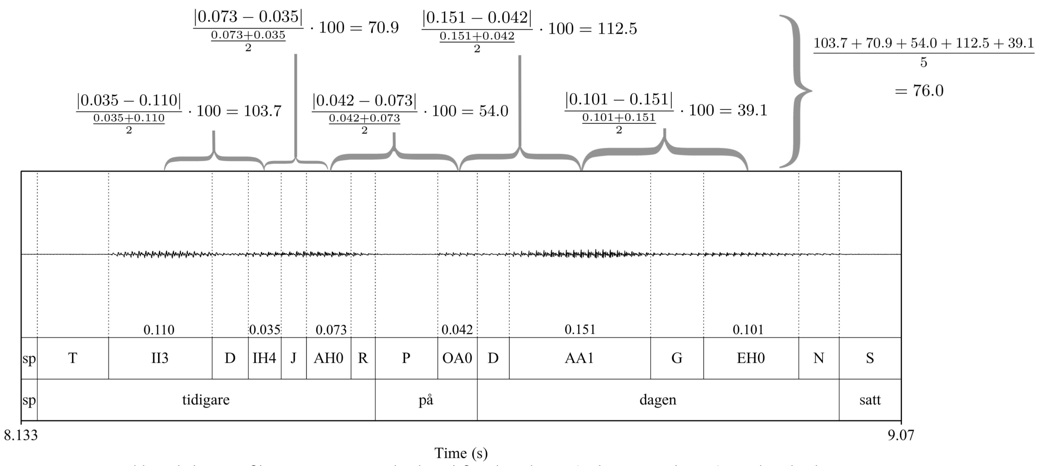


Figure 5.1: Visual breakdown of how nPVIV is calculated for the phrase 'tidigare på dagen'. Each calculation creates a tidy numeric representation for contrast: a larger number indexes high contrast and a low number indexes low contrast.

very different standard deviations. To normalize for rate, *variation coefficients* (the standard deviation divided by the mean) were introduced for consonants by Dellwo (2006) (VarcoC) and for vowels (VarcoV) by White and Mattys (2007). But variation coefficients (and standard deviations) still do not capture the pivotal *local* component that governs alternation as a construction. Specifically, 100-50-100-50-100-50 renders the same standard deviation and Varco as 100-100-100-50-50-50, even though the former contains more contrast than the latter. Local metrics provide a sophisticated resolution to this problem.

The *normalized pairwise variability index of vowels* (nPVIV) is the algorithm I use in my analysis. Developed by Low, Grabe, and Nolan (2000)[6], it was first used in a cross-varietal analysis of rhythm: Singapore English versus British English. nPVIV creates a numerical representation for the durational alternation of consecutive vowels and normalizes for speech rate:

$$nPVI = \frac{|d_{n+1} - d_n|}{\frac{d_{n+1}+d_n}{2}} \cdot 100 \tag{5.2}$$

where d = duration of the first vowel in the pair
where d_{n+1} = duration of the second vowel in the pair

Take, for example, the Swedish word *dagen* (*the day*) in Figure 5.1. AA1 lasts 0.151 seconds, and EH0 lasts 0.101 seconds, rendering 39.1 as the pairwise variability index for those two vowels, which is relatively low. On the other hand, the contrast between OA0 in *på* and AA1 *dagen* renders an nPVIV of 112.5, which is quite high. The pairwise index is typically calculated for every vowel pair for a speaker, and then the mean or median is taken to render the nPVIV for that passage. In the case of Figure 5.1, the mean is 76.0.

[6] Low et al. (2000) called it *pairwise variability index of vowels* (PVI) because the normalization and vocalic aspects of the algorithm were presupposed. It was only later, after non-normalized and non-vocalic spinoffs emerged, that 'normalized' and 'vowel' were added to the title.

nPVIV is typically used as a durational algorithm. However, in its first usage, Low (1998) calculated it from both vocalic duration and the amplitude integral. The latter is calculated by multiplying the duration of each vowel by its mean intensity. The discussion of integrals comes into relevance later on when I explain how I decide to operationalize rhythm (p. 123).

As is the case for my own investigation, Low et al. (2000:377) were also motivated by a sizeable body of literature that described Singapore English as 'staccato' or 'machine-gun' (Brown 1988; Platt & Weber 1980; Tay 1982; Tongue 1974). They found that Singapore English had lower overall nPVI than British English The authors also argue that nPVI is a superior metric to Ramus et al.'s %V, ΔV, and ΔC because the latter rendered the same results for Singapore English and British English. They argue that if the metrics were in fact reliable, Singapore English would have resulted in measurements closer to the Japanese side of Table 5.4. The problem, however, with their argument is that it relies on somewhat circular logic. The use of a mathematical metric should be motivated by how well it models a *prima facie* definition of rhythm instead of how well its results fit presumed rhythmic categories for languages. In Section 3, I therefore explicitly define rhythm for the purposes of this dissertation and then use this to motivate my selection of the appropriate mathematical metric.

nPVIV has been used in a number of studies. Grabe and Low (2002) examined one speaker of Thai, Dutch, German, British English, Tamil, Malay, SE, Greek, Welsh, Rumanian, Polish, Estonian, Catalan, French, Japanese, Luxembourgish, Spanish, and Mandarin each. They rank in nPVIV, from highest to lowest, according to the order that I listed them in. Thai was at 65.8, and Mandarin at 27.0 (Grabe & Low 2002:526).

In one of the largest comparative studies of speech rhythm to date, Thomas and Carter (2006) analyzed Spanish ($n = 8$), African-American ex-slave English ($n = 9$), Civil-war-era European-American English ($n = 5$), African-American English ($n = 20$), European-American English ($n = 20$), Gullah English ($n = 1$), Hispanic English ($n = 7$), and Jamaican English ($n = 4$). Two-hundred nPVIV quotients were examined from the 74 speech samples. Spanish showed the lowest alternation. The ex-slave, Jamaican, and Hispanic Englishes clustered in a middle range of alternation. The African-American, European-American, and Civil-war-era European-American Englishes clustered in the highest range of alternation (Thomas & Carter 2006:345).

Coggshall (2008) combined her corpus with Thomas and Carter's (2006) and compared the speech of 14 Cherokee-English and 25 Lumbee-English speakers with the speech of African-American and European-American speakers. She found that Cherokee-English and younger Lumbee-English speakers had lower alternation in their speech than African-American, European-American, and older Lumbee-English speakers.

Sarmah, Gogoi, and Wiltshire (2011) conducted an analysis of spontaneous and read-aloud Thai English ($n = 12$) and found that read-aloud speech was very similar in alternation to Grabe and Low's (2002) British-English speaker. In spontaneous speech, however, veteran speakers of Thai English had much lower alternation whereas new speakers of Thai English had alternation rates similar to read-aloud speech. They proposed that the latter were orient-ing toward an external norm (2002:86).

Torgersen and Szakay (2012) investigated contemporary London varieties on a large scale with a particular focus on multiethnolect, Multicultural London English (*MLE*). They found that young multiethnolectal speakers from Hackney ($n = 21$) had an overall lower nPVIV, followed by older speakers from Hackney ($n = 21$). The highest nPVIV values were for the remaining white speakers from both Hackney and Havering ($n = 40$). Not only do the findings show lower alternation in multiethnolect, they imply that the process started with Anglo residents of Hackney a full generation earlier (2012:829).

Shousterman (2015) examined spontaneous speech from 28 male and female speakers of New York Puerto-Rican English and found that younger speakers had *higher* alternation than older speakers (2015:171). She compared these results with Thomas and Carter's (2006) results and hypothesized that this was due to increased contact with African-American English among the younger speakers (2015:172).

Fuchs (2013) is unique because he uses more than just duration in his operationalization of nPVIV. He examined both spontaneous and read-aloud speech from educated Indian-English speakers ($n = 20$) and British-English speakers ($n = 10$). He extracted a battery of properties from the vowels in his dataset and used them to calculate various nPVIV indices. nPVIV of duration, intensity, and five measurements of loudness showed that educated Indian English had significantly lower rhythmic alternation than British English. nPVIV of sonority, voicing, and fundamental frequency rendered no significant differences. Further, spontaneous speech always had lower alternation than its read-aloud counterpart, ranging between 3.0 and 3.2 lower in durational nPVIV. None of the other non-nPVIV metrics, except White and Mattys' (2007) VarcoV, rendered significant differences in spontaneous speech (2013:98).

Fuchs (2013), like Low (1998), is also unique in that he incorporated *integrals* into his calculations of nPVIV – i.e., combined calculations of more than one vowel property. He calculated a duration-f_0 integral (duration · mean f_0) and an amplitude integral (duration · mean intensity) . While the nPVI of the duration-f_0 integral did not show a significant difference between Educated Indian English and British English, the nPVI of the amplitude integral did (2013:127, 129). In Section 3.3 (p. 123), I discuss the use of a similar integral for operationalizing rhythm in this dissertation.

Arvaniti (2009), however, has criticized the over-reliance on metrics. She found them inadequate for explaining classification differences between English, German, Greek, Italian, Korean, and Spanish.

> Rhythm classifications based on pooled data were inconsistent across metrics, while cross-linguistic differences in scores were often statistically non-significant even for comparisons between prototypical languages like English and Spanish. Metrics showed substantial inter-speaker variation and proved very sensitive to elicitation method and syllable complexity, so that the size of both ef-

fects was large and often comparable to that of language. These results suggest that any cross-linguistic differences captured by metrics are not robust (Arvaniti 2012:351).

It is important, however, to note that this critique is mainly aimed at two things; (1) typological classifications based on rhythm and (2) the rhythmic variability between speech styles within single languages. Arviniti did in fact find a significant main effect of language on all metrics tested, including %V, ΔC, rPVI-C (see definition below), VarcoV, and nPVI-V.

Gibbon (2003) has also taken issue with nPVIV. However, his critique was aimed at the mathematical premise of nPVIV. The algorithm assumes a strictly binary interpretation of rhythm and does not have the ability to model unary, ternary dactylic, or anapæstic rhythms. This can therefore result in an averaging out of otherwise important differences.

nPVIV studies in Swedish Three studies circulate to date that examine speech rhythm in Swedish with nPVIV. Krull (2012) analyzed the speech of women and showed that standard Central Swedish has an nPVIV that ranges between 46.0 and 57.4 (n = 5) and that L2 Swedish spoken by L1 Estonians has lower alternation, ranging between 44.7 and 53.0 (n = 5). Asu et al. (2017) analyzed recordings of read-aloud speech from three speakers of the Nücko variety of Estonian Swedish from 1920. They compared them with three elderly Central-Swedish speakers and found no significant inter-varietal difference for nPVIV (2017:76). The only significant difference was the durational variation of prosodic feet; Central Swedish had less variation while Nücko had more. In my earlier work, I examined the connection between speech rhythm of men and perceptions in Stockholm (Young 2018c). I found that men perceived as multiethnolectal had an nPVIV between 39.2 and 47.3 while men perceived as normative ranged between 52.6 and 54.9 (Young 2018c:50). In the pilot that led up to this study (my MA thesis), I also found that the distribution of nPVIV matched listener perceptions much more closely than Ramus et al.'s (1999) %V, ΔV, ΔC, VarcoV, and VarcoC (Young 2014:27–32).

Derivations of nPVIV Several other metrics circulate that are derivations of nPVIV. The *raw pairwise variability index* (rPVI) removed the normalization component and was used for consonants in Grabe and Low's (2002) comparative study of Thai, Dutch, German, British English, Tamil, Malay, SE, Greek, Welsh, Rumanian, Polish, Estonian, Catalan, French, Japanese, Luxembourgish, Spanish, and Mandarin. The rationale for removing the normalization component was never explicitly provided, but one could presume the assumption was that consonants do not reduce in rapid speech the way vowels do. Gibbon and Gut (2001) proposed using the nPVI formula for syllabic durations, and entitled this the *Rhythm Ratio* (RR). They found that Ibibio had lower alternation than British English (2001).

Gut (2005) later investigated rhythm in British English, Nigerian English, Hausa, Igbo and Yoruba. She found that Igbo had the lowest alternation, followed by Yoruba and Hausa in a tie, Nigerian English, then British English (2005:167). Deterding (2001) changed the normalizing denominator of nPVI to the overall durational mean and entitled it the *Variability Index* (VI). He used it on syllables to investigate rhythmic differences between British English and Singapore English. Like Low et al. (2000), he also found Singapore English to have significantly lower rhythmic alternation than British English.

3 Defining and operationalizing rhythm

3.1 Defining rhythm

Very few of the studies on rhythm actually specify how they define rhythm. White and Mattys (2007) argue that rhythm "derives from the repetition of elements perceived as similar" (2007:501). Arvaniti (2009) reminds us that White and Mattys' (2007) definition is derived from psychological research (Fraisse 1963, 1982; Woodrow 1951) and notes that

> the grouping of stimuli relies not just on duration but on a host of factors, including relative intensity, relative and absolute duration and the temporal spacing of elements (Fraisse 1963, 1982; Woodrow 1951). This definition of rhythm implies the presence of meter, which is distinguished from grouping itself: while grouping deals with phenomena that extend over time, meter is an abstract representation that relies on the alternation of strong and weak elements, not on absolute or relative durations (Lerdahl & Jackendoff 1983). (Arvaniti 2009:57)

Lerdahl and Jackendoff's (1983) definition of meter, cited above, is especially appealing because it pivots on the notion of salience. Rather than specifying an internal property, such as *long* versus *short* or *loud* versus *soft*, it focuses on the relativity component: *strong* and *weak*. White, Mattys, and Wiget (2012) adopt a similar definition for what they call *contrastive rhythm*, a feature that is "evident in any string of sounds in which there is an alternation of strong and weak elements" (2012:665). In this dissertation, I treat *meter* and *contrastive rhythm* as the same construction, namely, *rhythm*.

If we are to return to the many lay ascriptions of 'staccato' and investigate the truth behind them, we ought also ask how music scholars have defined 'staccato'. This field is relatively sparse, but a few moderately-cited studies circulate. Gabrielsson (1982) defines it by durational means whereby the duration $note_n$ is shorter than the inter-onset interval (IOI). The IOI is defined as duration between start of $note_n$ and the start of $note_{n+1}$. In this definition of 'staccato', large amounts of silence between each note appear to be of prime importance (Gabrielsson 1982:524). Bresin and Battel's (2000) study of five piano players enriched this definition by discovering that a key difference between staccato and legato is that for staccato, the IOI remained approximately twice the length of the note, even as the note fluctuated in duration (2000:219). Repp (1998) demonstrated this as well in his study of seven pianists: for staccato, "the adjusted note durations increased strongly with IOI" (1998:19). Not only do these findings implicate low alternation as a key feature of staccato, they exclusively operationalize staccato as a durational construction.

Durational operationalizations of rhythm dominate the field of phonetics as well. With the exception of Low (1998) and Fuchs (2013), every study I have cited in the prior sections conceptualizes and operationalizes rhythm as durational, so I will identify a few important exceptions here. In her 2010 dissertation, Cumming (2010) found that dynamic f_0 contributes to perceiving non-speech sounds and isolated monosyllables as longer than those without a dynamic f_0. However, she also found that listeners were more likely to assess stimuli as rhythmic when they had concordance between f_0 excursion and duration (2010:191). An open question is what this means from a production standpoint – is the latter assessment due to the fact that segments typically must be longer in order to operationalize any sort of suprasegmental excursion? I am unaware of any production studies that examined f_0 in this manner but, as I reviewed earlier, Low (1998) examined amplitude in her dissertation that prefaced Low et al.'s (2000) seminal study. She calculated the nPVI for root mean square (RMS) amplitude and found higher RMS amplitude nPVI values for British English than for Singapore English (1998:52–53). These two studies offer promising enrichments to a field that otherwise has primarily conceptualized rhythm as a durational construct.

3.2 Defining prominence

In order to operationalize Lerdahl and Jackendoff's (1983) 'strong' and 'weak', it becomes necessary to discuss and define prominence. There is general agreement that prominence consists of the following three cues: duration, fundamental frequency, and intensity (Breen, Fedorenko, Wagner, & Gibson 2010; Fry 1955, 1958; Kochanski, Grabe, Coleman, & Rosner 2005; Lieberman 1960; Turk & Sawusch 1996; Wagner & Watson 2010). Researchers, however, disagree on the extent to which each cue contributes to the construction, and this is complicated by the fact that it may differ by language (Wagner & Watson 2010:925).

Turk and Sawusch (1996) found that duration played the most important role in the perception of English prominence whereas intensity played little to no role. Forty listener participants assessed a series of guises for the word 'mama' in which the fundamental frequency had been neutralized and duration and intensity were altered, rendering four possible stimuli: short/soft, short/loud, long/soft, long/loud. Half the speakers were asked to assess whether they heard a loud first syllable and the other half were asked to asses whether they heard a long first syllable. Listeners showed more interference from irrelevant variation in duration when attending to loudness than they did from variation in intensity when attending to duration (1996:3785). The authors then conducted a second experiment in which stimuli containing máma and mamá were assessed by 66 listeners for quality on a six-point scale. Again, the fundamental frequency was normalized while intensity and duration were altered at five different levels each and combined to produce 25 possible combinations. For each set, five duration values were orthogonally combined with each of five intensity values to produce 25 stimuli. They found that "computed measures of length consistently predicted prominence judgments more accurately than did loudness" (1996:3789).

On the other hand, Kochanski et al. (2005) found loudness played the biggest role in prominence detection in the perception of prominence for seven British English urban dialects. Duration played a secondary role, and fundamental frequency played little to no role. They investigated the correlation between approximately 2000 professionally-assessed prominence marks and perceptual loudness, phone duration, aperiodicity, spectral slope, and fundamental frequency. While loudness carried the strongest correlation – followed by duration – aperiodicity, spectral slope and fundamental frequency were mostly orthogonal to the assessments of prominence (2005:1046).

Breen et al. (2010) found that intensity and duration played consistent roles in the perception of all types of prominence for American English, including general focus location, narrow object focus, and contrastive focus. Fundamental frequency operated differently, depending on the type of focus. General focus and narrow object focus were perceived most accurately with high fundamental frequency, and contrastive focus with low fundamental frequency (2010:1044).

Studies on English prominence have typically tried to explain what distinguishes them from their non-prominent counterparts. Studies on Swedish prominence have typically tried to explain what distinguishes its two prominence categories from each other: accent 1 from accent 2. Because pitch accent is typologically quite marked, this is obviously a much more exciting question. But it has inadvertently resulted in less work on the distinction between prominence and lack of prominence in Swedish. In much of the work on pitch accent, the implicit assumption is that fundamental frequency is the most pivotal element of Swedish prominence (Engstrand 1995, 1997; Myrberg 2010; Riad 1998, 2006, 2009).

	SMALL accent (*usually ω stress*)	BIG accent (*usually φ head*)
Accent 1	H L*	L* H
Accent 2	H* L	H*L H

Table 5.5: Description of pitch accents for Stockholm Swedish as described by Myrberg and Riad (2015:116): SMALL and BIG accents 1, 2, and compound accent 2. Examples are shown with orthodox contours. *ω* stands for word accent, and *φ* stands for phrase accent.

As I discuss on pages 45–47, Central Swedish, to which Stockholm Swedish belongs, has two lexical pitch accents that are distinguished by the quality and timing of fundamental frequency. Further, they manifest themselves differently, depending on whether they are SMALL or BIG accents. Myrberg (2010) has argued that the former accent often denotes general focus while the latter contrastive. Later assessments claim that the relationship between form and function is too far from straightforward to settle on this definition (Myrberg & Riad 2015). Therefore, four tonal configurations are said to mark prominence, and fundamental frequency has been identified as the most important distinguishing feature (Engstrand 1995, 1997; Myrberg 2010; Riad 1998, 2006, 2009). Table 5.5 shows the four tonal configurations.

Turning to the construction of prominence in relation to the construction of non-prominence in Swedish, Fant and Kruckenberg (1994) found duration to be the most important correlate and note that this is unsurprising since Swedish is a quantity language (see my review of phonological quantity on pp. 43–45). However, they note also that f_0 is of near-equal importance.

> In our view, duration is the most consistent physical correlate of stress. Next, or of equal importance, is the f_0 pattern of the associated word accent. However, we find examples of weak stress realised by a significant duration increase but without f_0 modulation. On the other hand, there are also examples of focal accent with appreciable f_0-modulation but without appreciable duration or intensity increase. (Fant & Kruckenberg 1994:141)

Heldner and Strangert (2001) similarly found that focus (BIG) accents involve durational lengthening that extends to the subsequent unstressed syllable. Any additional unstressed syllables to the right of the first unstressed syllable did not lengthen (2001).

Fant and Kruckenberg (1994) also observed that intensity typically correlates highly with f_0 except at the very peak of an f_0 swing, whereby there often appears to be an inverse relationship between energy and f_0 (1994:137–141). In later work, they expanded their study to perception and tested the correlation between assessments of prominence with duration, f_0, energy, and subglottal pressure (Fant, Kruckenberg, & Liljencrants 2000). Listeners were asked to assess syllables in a series of spoken stimuli by giving each syllable a prominence rating between 0 and 30. The rating is referred to by the authors as *Rs*. Subjects were told that typical values for unstressed syllables would be $Rs = 10$ and typical values for stressed syllables would be $Rs = 20$.

> In summary, an increase of Rs of an accented syllable by 10 units from 15 to 25 is associated with 125 ms increase of duration, about 4-8 semitones in f_0, 6 dB in sound pressure level and 9 dB in sound pressure level with high frequency preemphasis.(Fant et al. 2000:81)

The authors find that all three parameters have correlations with prominence in Swedish. They caution, however, that this does not mean all components weigh equally in terms of perception. They suggest that future studies ought to attempt to remove single elements from prominent syllables in order to assess relative salience (Fant et al. 2000:82).

Strangert and Heldner (1995) asked nine phonetically-experienced transcribers to rate 60 accent-1 and 55 accent-2 words for prominence. They found that the "the greater the f_0-rise, the stronger the agreement on focus accent. That is, the size of the focus accent cues the degree of prominence" (1995:59). They note, however, that it only partially explains the variation among the assessments. The correlation between semitones and prominence ratings was of intermediate strength with R^2 values of 0.14 for read-aloud accent 1, 0.41 for read-aloud accent 2, 0.26 for spontaneous accent 1, and 0.26 for spontaneous accent 2 (1995:59).

Heldner and Strangert (1997) later found that "f_0-rise is neither necessary nor sufficient for the perception of focus." (1997:55). This was concluded after playing a series of guises to listeners in which f_0 contours on prominent syllables were artificially flattened. Listeners were still able to assess them as prominent. In subsequent work, Heldner (2003) found that overall intensity and spectral emphasis were strong correlates to focus (BIG) accents in production. He notes, however, that "the reliability of acoustic correlates is not the same as the reliability of perceptual cues" (2003:57).

3.3 Operationalizing rhythm

In light of the literature reviewed above, this study will incorporate duration, intensity, and fundamental frequency in its operationalization of rhythm by using an adaptation of Mishra et al.'s (2012) *energy-f_0 integral* (EFI). EFI was developed for the automatic detection of prominence. When the authors tested it against a manually-annotated English-language corpus of 67 000 words, they found EFI to reliably detect the words that had been annotated as prominent (2012:4). Mehrabani, Mishra, and Conkie (2013) successfully used EFI in their computational model for detecting *Just-Noticeable-Differences* in pitch, duration, and energy.

Rhythm will be calculated using Low et al.'s (2000) nPVIV algorithm, and the measurements for the vowels will be calculated using Mishra et al.'s (2012) EFI.

$$EFI = d_k \cdot energy_k \cdot f_{0k} \tag{5.3}$$

where d_k = duration of segment at k^{th} position
where $energy_k$ = mean intensity of segment at k^{th} position
where f_{0k} = mean fundamental frequency of segment at k^{th} position

$$nPVIV_{EFI} = \frac{|EFI_{n+1} - EFI_n|}{\frac{EFI_{n+1}+EFI_n}{2}} \cdot 100 \tag{5.4}$$

where EFI_n = Energy-f_0-integral of the first vowel in the pair
where EFI_{n+1} = Energy-f_0-integral of the second vowel in the pair

	$nPVIV_{energy}$	$nPVIV_{f_0}$	$nPVIV_{EFI}$	$nPVIV_{duration}$
$nPVIV_{energy}$	1.00			
$nPVIV_{f_0}$	0.20	1.00		
$nPVIV_{EFI}$	0.24	0.40	1.00	
$nPVIV_{duration}$	0.10	0.03	0.79	1.00

Table 5.6: Correlation matrix of nPVIV calculated with energy, f_0, EFI, and duration.

Importantly, this is not a proposal for yet another rhythm metric (there are already too many in circulation). Fuchs (2013) also used integrals in his calculation of nPVIV. As I reviewed earlier, he calculated a duration-f_0 integral (duration · mean f_0) and an amplitude integral (duration ·x mean intensity). In neither case were these conceptualized as novel metrics. Rather, they represent a small adjustment to an already well-established rhythm metric, the nPVIV.

Table 5.6 shows a correlation table for the three components of this new calculation – mean intensity, mean f_0, and duration – along with the new EFI calculation. First, it demonstrates that the three components of EFI are relatively orthogonal to one another. This is important because it means that the EFI index is not merely a proxy that exponentializes the effects of duration. Second, it shows that $nPVIV_{duration}$ contributes the most to the contrastive element in the $nPVIV_{EFI}$ calculation, followed by $nPVIV_{f_0}$, and then followed by $nPVIV_{intensity}$.

One might argue that Table 5.6 is justification for not combining these parameters together into a single metric. For example, it might obscure the investigation of segmental transformations in multiethnolect. However, the main research question is whether rhythm in Stockholm's multiethnolect is *staccato* rather than whether multiethnolect has differences in quantity (or f_0 or intensity). The former is a question of a latent construction; the latter are questions of descriptive phonetics.

Since the latent construction of prominence in Swedish relies on all three parameters in production, the operationalization of rhythm (i.e., *strong-weak*) ought to include all three. If variation in this construction is found to correlate with social parameters, the next question is then what phonetic transformations lie behind this characteristic. Chapters 6 and 8 examine the question of how rhythm as a latent construction varies across social and contextual parameters. Chapters 7 and 9 examine what phonetic components lie behind this variation.

Although my calculation differs from the established norm, I see no need to use a new term like $nPVIV_{EFI}$ throughout the dissertation. This is because the term *normalized pairwise variablity index of vowels* does not specify what measurements are taken. Rather, it just specifies where the measurements are taken; namely, vowels. Therefore, I will continue to use the unembellished designation *nPVIV*.

It is important to note that operationalizing rhythm with EFI – while a more accurate way to numerically represent prominence alternation – renders metrics that are not cross-comparable. Since the preceding literature has relied on nPVIV of duration, the results from this study cannot be used by future scholars who wish to see where Swedish rhythm places in respect to other languages. Nonetheless, as I have stated before, the aim of this investigation is to operationalize rhythm in a way that ties closely to how rhythm is in fact constructed. This is a mandate borne out of local observations of "staccato" and the coinciding global literature that has shown that duration is just one of many components behind the aggregated construction of prominence. I am therefore of the opinion that any numerical operationalization must model the coinciding aggregations of intensity, f_0 and duration.

The cross-comparative studies often beg the question of means (or medians), which is something that this dissertation has decidedly departed from. As I demonstrate in the next chapter, each pairwise contrast is highly variable within a single language. This is because the individual pairs are affected by all sorts of internal constraints – such a prominence or the lack thereof – as well as mundane methodological decisions such as including rhotic codas within the vocalic measurement as in Thomas and Carter (2006:341) and this study or excluding them as in Low et al. (2000). What I mean to say is that since cross-study comparisons are questionable endeavors even when the same metric is used, the opportunity cost of using a different metric is less high.

Citations in original language

[i] Även i uttalet förekommer en del drag som skulle kunna betecknas som förenklande, t ex en tendens hos många talare att reducera skillnaden lång – kort stavelse. Tillsammans med vissa prosodiska fenomen och ibland också en tydligare markering av ordgränser än normalt gör detta att infödda svenskar ibland upplever talet som 'stötigt'. (Kotsinas 1988a:268)

[ii] Det mest påfallande draget hos flera av talarna, speciellt A3, D5 och F4, är emellertid något som av lyssnare brukar karakteriseras som 'stötighet'. Exakt vad denna stötighet beror på är svårt att avgöra utan en ingående fonetisk analys. (Kotsinas 1990:257)

[iii] De viktigaste dragen som utmärker hjällbospråket gäller prosodin. Det sätt man talar på karakteriseras ibland som "stötigt". Det innebär att fördelningen av betoning på både ord- och satsnivån skiljer sig från standardspråket, tonaccenter omfördelas eller upphävs och vokallängder förkortas. Dessa drag kan tillsammans ge ett främmande, ibland entonigt intryck. (Nordenstam & Wallin 2002:145)

Chapter 6

RHYTHM IN STOCKHOLM'S VERNACULAR

In the preceding chapter, I identified the literature that has characterized Stockholm's multiethnolect as 'staccato'. I then contextualized this literature within a more global genre that has found contact varieties to have rhythmic features that are compromises between their respective superstrates and substrates. This literature has also identified preferred approaches for operationalizing rhythm, including the *normalized pairwise variability index of vowels* (*nPVIV*, Low et al. 2000).

In this chapter, I will examine a large dataset of Stockholm Swedish that contains the vernacular speech of 36 male speakers, ages 24–43, from a representative sample of social-class and ethnic backgrounds. The linguistic variable is rhythm, and I operationalize it with nPVIV. As I outlined in the previous chapter, the vowel property used in the formula will not be duration, but rather, Mishra et al.'s (2012) *energy-*f_0 *integral* (*EFI*).

1 Research question and experimental design

As was reviewed in Chapter 5, there is substantial literature on the impression of 'staccato' rhythm in Swedish multiethnolect (Kotsinas 1988a:268; Kotsinas 1990:257; Kotsinas 1994b; Kotsinas 2000; Nordenstam and Wallin 2002:145; Fraurud 2003:87; Bijvoet and Fraurud 2008; Bijvoet and Fraurud 2011:16; Bodén 2011; Milani and Jonsson 2012:54; Young 2018c). There is also substantial literature on low alternation in European multiethnolects (Duez and Casanova 2000; Cerquiglini 2001; Fagyal 2010:117; Hansen and Pharao 2010:91; Fagyal and Stewart 2011; Torgersen and Szakay 2012; Kern 2013; among others). Therefore,

> RQ: *Does the speech of Stockholm's racialized working class have lower rhythmic alternation than other varieties in the city; i.e., is it 'staccato'?*

Three ensuing analyses will address this research question.

Design and roadmap

Section 2 Modeling internal factors   I demonstrated in my review of the literature that rhythm is an abstract phenomenon that at the very least is influenced by multiple phonological features and at the very most is the byproduct of multiple phonological features. Therefore, no examination of external social predictors should be conducted without first identifying language-internal constraints. This analysis will identify potential phonological/internal predictors of nPVIV and test their combined simultaneous effects on nPVIV.

Section 3 Introducing class-based parameters to the model   I will add social predictors to the base statistical model from section 2 in order to test their significance and effects on model fit. Because the literature has near universal consensus in ascribing multiethnolect to young, racialized, and working-class speakers, the predictors selected for this analysis are age, race, and social class. For social class, six parameters are tested in terms of their effects on nPVIV: socioeconomic index (SEI), income, formal education, parental formal education, parental SEI, and taste.

Section 4 Introducing social-network parameters to the model   I quantitatively operationalize qualities of the speakers' social networks and add them to base statistical model from section 2. Network predictors total six in number – three predictors are proxy network calculations, and three predictors are actual network calculations. For proxy networks, I examine (1) the percentage of working-class classmates in the schools that the participant attended as a child, (2) the percentage of classmates of 'foreign descent' in the schools that the participant attended as a child, and (3) the average of the percentage of residents of 'foreign descent' of every neighborhood the participant has lived in. For actual networks, I examine (1) the percentage of close contacts to the participant who speak 'hood' (Sw.: 'orten'), (2) the percentage of close contacts to the participant who are racialized and working-class, and (3) the percentage of close contacts to the participant who are white and working-class. These six parameters are tested in terms of their predictive effects on nPVIV.

2 Modeling internal factors

Rhythm is an abstract phenomenon that at the very least is influenced by multiple phonological features and at the very most is the byproduct of multiple phonological features. Therefore, no examination of external social predictors can be conducted without first identifying language-internal constraints. This analysis will identify potential phonological/internal predictors of nPVIV and test their combined simultaneous effects on nPVIV.

According to Tagliamonte (2006), any investigation of sociolinguistic variation must first factor in internal predictors (2006:104–205). For example, in Sharma and Sankaran's (2011) analysis of (t) in a London Asian community, the authors included preceding segments, following segments, and word class as potential predictors. These were modeled alongside gender, age, formality, and so on (2011:412).

Studies of speech rhythm have been the puzzling exception. To date, no study has modeled nPVIV at the observation level while taking internal predictors into account. The typical procedure is to model social predictors against the speaker's mean or median nPVIV while including mean speech rate as the sole internal parameter (see, e.g., Torgersen and Szakay 2012:830). I argue that we overlook important variation when we exclude pair-by-pair measurements in our statistical models of rhythm. My argument is exemplified in Figure 6.1 where intra-speaker nPVIV moves from 103.7 to 70.9 to 54.0 to 112.5 to 39.1 for just a single sequence of vowels. Clearly, the pair-by-pair nPVI calculations are highly varied and governed by vigorous phonological factors[1]. Such internal factors must certainly restrict to what extent a speech community can commandeer rhythm for social meaning.

Tables 6.1 and 6.2 offer a visualization of how this study differs in its analytic approach. Table 6.1 shows how my data would look in a traditional dataset modeled on means/medians. Table 6.2 shows my data in the 'new' approach.

This first analysis will only model internal predictors. The analyses in Sections 3 and 4 will investigate the predictive power of social predictors by adding them to the baseline internal model. In all models, random effects of speaker and vowel are taken into account.

2.1 Method

Extracting the data subset for CASUAL speech

This analysis includes the speech of all 36 speakers from the main dataset. As outlined in the Methods Chapter, speech data comes from the fieldwork I conducted between 2015 and 2018

[1]NB that Figure 6.1 shows nPVIV as a *durational* calculation whereas this study calculates nPVIV from the *duration* · F_0 · *intensity* of each vowel.

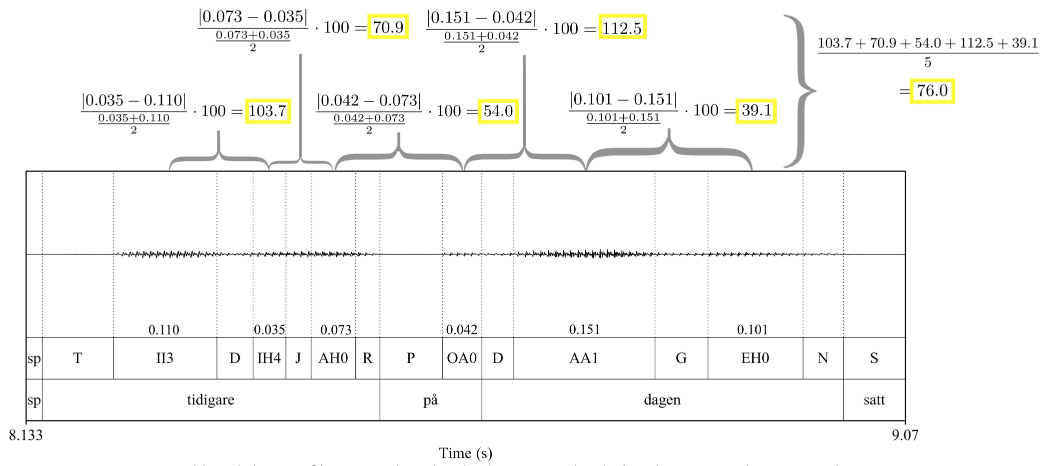


Figure 6.1: Visual breakdown of how much individual variation lies behind aggregated nPVIV values.

in Stockholm. Where peer-group conversations were available, data was extracted there. For speakers for whom I had no peer-group recording, spontaneous speech was extracted from the interview. Criteria for treatment as 'spontaneous' was the presence of swearing, *channel cues* (Labov 1972b:113) and/or a topic that was engaging for the speaker such as danger of death or supernatural occurrences (Labov 1972b).

The interviews and group conversations were transcribed by myself and a team of transcribers. The transcriptions were phonetically-timed aligned automatically via SweFA (Young 2018b; Young & McGarrah 2017). I manually corrected all segment boundaries after forced alignment. The extracted samples range between 295 and 1 517 vocalic elements per speaker. Some de facto sample-size benchmarks that prior studies have set range between 72 and 661 (Torgersen & Szakay 2012:828), between 163 and 588 (Fuchs 2013:98), and 200 vocalic intervals per speaker (Thomas & Carter 2006:340).

Phonetic extraction and annotation

Segmental metrics were extracted using a customized adaptation of Brato's (2015) script for Praat (Boersma & Weenink 2017), a copy of which is provided in Appendix B.

The final foot before a pause was included in the calculation (Torgersen & Szakay 2012, Fuchs 2013, White & Mattys 2007, Low et al. 2000, though cf. Thomas & Carter 2006, Sarmah et al. 2011).

I delineated breath groups by pauses that exceeded 150 milliseconds. This is in line with Fuchs (2013:92) but contradicts Thomas and Carter's (2006) recommendation of 70 milliseconds. I chose a higher number because Swedish geminates and consonant clusters often exceed 70 milliseconds (Engstrand & Krull 1994). If we so readily accept that an nPVIV contrast that straddles geminate consonants is meaningful, then an nPVIV contrast straddling pauses of an equivalent duration might also be meaningful[2].

[2]I concede, however, that there is always the possibility that pauses trigger some sort of cognitive process that

In Central Swedish, coda /r/ typically coalesces by means of a sandhi process by which the subsequent /d, n, s, t/ become [ɖ, ɳ, ʂ, ʈ], respectively (Riad 2014). In front of other consonants, it usually occurs as an approximant[3] or is partially or fully elided. Syllable-final /r/ was included as part of the vowel because the boundaries V+/r/ before C are highly subjective (Thomas & Carter 2006:341). Syllable-final /j/ was included as part of the preceding vowel for the same reason. The adjoinment of measurements of coda /j/ and /r/ was performed with a TextFactory script I wrote in *BBEdit*, provided in Appendix C.

On the other hand, intersyllabic /r/ and /j/ (V+/r j/+V), including the external-sandhi effect of coda /j, r/ + onset V (e.g., *han är ung* ⟶ *han ä r̲ung*), were treated as consonants.

Hesitation markers and hesitation lengthening were manually removed as I encountered them. As an additional safeguard, I removed all nPVIV values from the data that were greater than two times the standard deviation plus the mean of the entire dataset ($\mu + 2\sigma$).

2.2 Internal predictors and statistical model

I built a mixed-effects linear regression model with internal factors as predictors, nPVIV in CASUAL speech as the response variable, and speaker and vowel pair as random effects. The predictors are the presence of a phonologically-long accent-1 vowel (ACCENT_1_LONG), the presence of a phonologically-short accent-1 vowel (ACCENT_1_SHORT), the presence of a phonologically-long accent-2 vowel (ACCENT_2_LONG), the presence of a phonologically-short accent-2 vowel (ACCENT_2_SHORT), the presence of a phonologically-long unstressed vowel (UNSTRESSED_LONG), the presence of a phonologically-short unstressed vowel (UNSTRESSED_SHORT), the presence of a phrase-final vowel (PHRASE_FINAL), the presence of a rhotic in coda position (CODA_R), word frequency of the word(s) within which the pair occurs (LEXICAL_FREQ), and the speech rate within which the pair occurs (SPEECH_RATE).

In the below sections, I describe and motivate the inclusion and operationalization of each predictor.

Pitch accent, acoustic prominence, and phonological length

This section introduces and describes the following predictors:

coding ID: ACCENT_1_LONG
coding ID: ACCENT_1_SHORT
coding ID: ACCENT_2_LONG
coding ID: ACCENT_2_SHORT

As I describe in Chapter 3 Section 1.2 (p. 45), Swedish is a pitch-accent language with two lexical accents: *accent 1* and *accent 2*. As I describe in Chapter 3 Section 1.1, it also has two

attenuates the perception of proximity. That said, the assumption in this study is that there is *not*.

[3] This is not always the case in the data. Trills and taps do occasionally occur, which further necessitates accounting for any R-ful observation in the statistical model.

			Internal Factors	External Factors			Response
	STYLE	SPEAKER	MEAN_SPEECH_RATE	SEI_ADULT	RACIALIZATION	AGE	MEDIAN_NPVI_V
1	READING	ABEL	228	51	INVANDRARE	26	48.8
2	READING	AUGUST	262	73	SVENSK	42	54.4
3	READING	ANTONIO	257	11	INVANDRARE	24	45.1
			lines 4 through 105				
106	CASUAL	SHORTY	195	53	INVANDRARE	26	45.4
107	CASUAL	TARIK	217	51	INVANDRARE	31	42.6
108	CASUAL	THOMAS	202	57	SVENSK	32	44.9

Table 6.1: Excerpt of the dataset that a traditional nPVI analysis would use. There would be 108 observations (36 speakers · 3 styles), one internal predictor of mean speech rate, the 14 external (social) predictors, and a response variable of median nPVIV.

		Random effects		Internal Factors										External Factors			Response
	STYLE	(1|SPEAKER)	(1|VOWEL)	ACCENT_1_LONG	ACCENT_1_SHORT	ACCENT_2_LONG	ACCENT_2_SHORT	UNSTRESSED_LONG	UNSTRESSED_SHORT	PHRASE_FINAL	CODA_R	SPEECH_RATE	LEXICAL_FREQ	RACIALIZED_BINARY	AGE	SEI	NPVIV
1	READING	ABEL	IH1_RR_UH0	NO	YES	NO	NO	NO	YES	NO	YES	238	3	INVANDRARE	26	51	90.4
2	READING	ABEL	UH0_EH0	NO	NO	NO	NO	NO	YES	NO	NO	156	3	INVANDRARE	26	51	37.3
3	READING	ABEL	EH0_AA0	NO	NO	NO	NO	YES	YES	NO	NO	163	3375	INVANDRARE	26	51	23.7
4	READING	ABEL	AA0_OAH0	NO	NO	NO	NO	YES	YES	NO	NO	160	3401	INVANDRARE	26	51	59.9
5	READING	ABEL	OAH0_AE1	YES	NO	NO	NO	NO	YES	YES	NO	225	54	INVANDRARE	26	51	105.4
6	READING	ABEL	EE3_AH4	NO	NO	NO	NO	YES	YES	NO	NO	151	244	INVANDRARE	26	51	33.3
7	READING	ABEL	AH4_AH0	NO	NO	NO	NO	NO	YES	NO	NO	236	123	INVANDRARE	26	51	19.8
8	READING	ABEL	AH0_AA1	YES	NO	NO	NO	NO	YES	NO	NO	309	1	INVANDRARE	26	51	123.8
9	READING	ABEL	AA1_AEH1	YES	NO	NO	NO	NO	YES	NO	NO	320	3657	INVANDRARE	26	51	115.7
							lines 10 through 43003										
40 269	CASUAL	TARIK	EH1_AEH0	NO	NO	NO	NO	NO	YES	NO	NO	188	545	INVANDRARE	31	51	39.3
40 270	CASUAL	TARIK	AEH0_IH3	NO	NO	NO	YES	NO	YES	NO	NO	204	9708	INVANDRARE	31	51	64.1
40 271	CASUAL	TARIK	EH4_AH1	NO	YES	NO	NO	NO	YES	NO	NO	203	19605	INVANDRARE	31	51	59.1
40 272	CASUAL	TARIK	EH1_AE1	YES	NO	NO	NO	NO	YES	NO	NO	237	1698	INVANDRARE	31	51	43.3
40 273	CASUAL	TARIK	AE1_EH1	YES	NO	NO	NO	NO	YES	YES	NO	293	4519	INVANDRARE	31	51	46.7
40 274	CASUAL	TARIK	UH0_EE2	NO	NO	NO	NO	YES	YES	NO	NO	225	1	INVANDRARE	31	51	96.8
40 275	CASUAL	TARIK	EE2_AH4	NO	NO	NO	NO	YES	YES	NO	NO	167	1	INVANDRARE	31	51	39.6
40 276	CASUAL	TARIK	AH4_EH0	NO	NO	NO	NO	NO	YES	NO	NO	180	1	INVANDRARE	31	51	46.8
40 277	CASUAL	TARIK	EH0_YH0	NO	NO	NO	NO	NO	YES	NO	NO	182	41	INVANDRARE	31	51	34.9

Table 6.2: Excerpt of the dataset that this analysis uses. There are 40 277 total observations (36 speakers · 3 styles · [295 to 1517 observations per speaker]), of which 21 000 are CASUAL speech. The two random intercepts are speaker and vowel combination. The nine internal predictors are accent 1 long, accent 1 short, accent 2 long, accent 2 short, unstressed long, unstressed short, phrase-final syllable, coda /r/, mean frequency of constituent lexemes, and speech rate. The three external social predictors are race, SEI, and age. The response variable is observation-level nPVIV. Note that since this is just the data-set used to run the statistical models, the calculations of prominence or the separations between intonational phrases are not included here.

speaker	vowel	duration	mean dB	mean f_0	EFI calculation	EFI	start time	end time	time-distance-weighting calculation	time-distance weighting	EFI*weighting	
ABEL	UH0	0.05	51	138	$0.047 \cdot 51 \cdot 138$	328		64.26	$e^{-1.6 \cdot (64.53-64.26)}$	0.65	213	
ABEL	IH0	0.03	52	137	$0.031 \cdot 52 \cdot 137$	219		64.36	$e^{-1.6 \cdot (64.53-64.36)}$	0.77	168	
ABEL	EH0	0.04	52	129	$0.04 \cdot 52 \cdot 129$	268		64.44	$e^{-1.6 \cdot (64.53-64.44)}$	0.87	232	prominence calculation
ABEL	OEE6	0.1	55	130	$0.097 \cdot 55 \cdot 130$	703	64.53	64.63				if 703 > {213, 168, 232, 399, 196, 125}, then vowel is PROMINENT
ABEL	EH7	0.06	52	128	$0.065 \cdot 52 \cdot 128$	434	64.68		$e^{-1.6 \cdot (64.68-64.63)}$	0.92	399	
ABEL	EH0	0.04	52	126	$0.045 \cdot 52 \cdot 126$	298	64.89		$e^{-1.6 \cdot (64.89-64.63)}$	0.66	196	
ABEL	EH0	0.04	50	127	$0.037 \cdot 50 \cdot 127$	234	65.02		$e^{-1.6 \cdot (65.02-64.63)}$	0.53	125	

Table 6.3: A breakdown of how prominence is calculated and coded for the vowel OEE6 in *cirkusdirektören*. The energy-f_0-integral (EFI) of the surrounding vowels are time-distance weighted according to their distance from OEE6. If OEE6 *still* has the largest EFI after that weighting, then it is coded as YES. If not, it is coded as NO. NB: Although this table only shows the calculation for OEE6, the analysis performs the calculation for *all* vowels, including UH0, IH0, EH0, EH7, EH0, and EH0 above

phonological categories of vowels – *long* and *short* (p. 43). These interact to form four possible combinations, long vowels with accent 1 (ACCENT_1_LONG), short vowels with accent 1 (ACCENT_1_SHORT), long vowels with accent 2 (ACCENT_2_LONG), and short vowels with accent 2 (ACCENT_2_SHORT).

When examining rhythm via a blanket algorithm like nPVIV, the obvious connection between stress and rhythm means that pitch accent must be taken into account. This is because there is always the risk that the algorithm is simply capturing some single extraneous feature like pitch accent. Avoiding this risk is all the more critical when one takes into account the finding that Oslo's multiethnolect has been found to use the Norwegian pitch accent 1 in place of pitch accent 2 (Svendsen & Røyneland 2008:71–74).

Prominence is a highly subjective construct that is often coded manually and influenced by contextual clues. Recall in Section 1.2, however, that Myrberg and Riad (2015) argue that there is no straightforward connection between form and function for pitch accent in Swedish. Motivated by this grammatically neutral and highly acoustic definition, I code pitch accent according to the presence or absence of *relative acoustic prominence*. A vowel is coded as prominent if it has a higher EFI than the three preceding and subsequent vowels, adjusted by a sigmoidal time penalty $e^{-1.6t}$ where t = time between the boundaries of the two vowels at hand. The calculation was made using the following steps:

1. *Energy-f_0 integral* (*duration* $\cdot$ *f_0* $\cdot$ *intensity*) is calculated for the vowel in question and the three vowels *before* it and three vowels *after* it[4].

2. The time distance from the three vowels *before* it and the time distance from the three

[4] This is adapted obviously to when the speaker stops or begins speaking. So the first vowel in the data will only be compared to vowels after it. Likewise, the final vowel in the data is only compared with vowels before it.

vowels after it are calculated, rendering six distances.

3. Those six time distances t were incorporated into the sigmoid function $e^{-1.6t}$ to create a weighted index. I choose this factor by trialing various factors against my own subjective evaluations of prominence.

4. Each of those weights were then multiplied by their corresponding energy-f_0 integrals.

5. If the EFI of the vowel in question was larger than the six weighted EFI's, then it was coded as YES for prominent. If not, it was coded as NO.

6. This was calculated for every vowel.

Table 6.3 provides a detailed outline of the above steps for vowel OEE6 in *cirkusdirektören*. This same calculation was run for the remaining vowels in the data ($n = 40\,277$)

I settled on the sigmoidal penalty $e^{-1.6t}$ by means of a trial-and-error experiment that I ran on a small subset of the data. I coded a short section by hand for pitch accent – both small and large – and experimented with various sigmoid functions that would acoustically also mark these same spots as prominent. $e^{-1.6t}$ was the optimal formula. After applying the formula to the entire dataset, portions were extracted for spot checks. The annotation of prominence in these spot checks correlated with my own perceptual assessments.

Coding for prominence was integrated into the coding for lexical pitch accent and the coding for phonological length, rendering *accent 1 long*, *accent 1 short*, *accent 2 long*, and *accent 2 short*[5].

Table 6.2 offers an example for how the coding is distributed among the four accent/quantity combinations. If either vowel in the nPVIV pair is prominent, phonologically long, and part of an accent-1 word, then the predictor ACCENT_1_LONG is coded YES as is the case for the pairs that contain AE1 [^{1}ɛː] on lines 5, 40 272, and 40 273 and for the pairs that contain AA1 [^{1}ɑː] on lines 8 and 9. If either vowel in the nPVIV pair is prominent, phonologically short, and part of an accent-1 word, then the predictor ACCENT_1_SHORT is coded YES as is the case for the pair that contains IH1 [^{1}ɪ] on line 1 and AH1 [^{1}a] on line 40 271. If either vowel in the nPVIV pair is prominent, phonologically long, and part of an accent-2 word, then the predictor ACCENT_2_LONG is coded YES (no example available in Table 6.2). If either vowel in the nPVIV pair is prominent, phonologically short, and part of an accent-2 word, then the predictor ACCENT_2_LONG is coded YES as is the case for the pair that contains IH3 [^{2}ɪ] on line 40 270 in Table 6.2.

In prior reviews, there has been concern that the nPVIV of EFI is not sufficiently orthogonal to a prominence annotation that is based on relative EFI. I dispute this. The calculation of nPVIV backs out raw EFI values, rendering a simple percentage difference. The annotation of prominence described here identifies the relatively highest EFI – in binary YES-NO terms – to a pool of surrounding time-weighted EFI values. In this sense, it simulates manual

[5]NB, accent-2 compounds are annotated within the same category as regular accent-2 words.

coding of prominence, albeit in a more systematic fashion. In neither calculation do nominal EFI values survive. Therefore, I see no *a priori* collinearity between the prominence annotation here and the calculation of nPVIV. Table 6.4 provides a list of every possible vowel-pair combination and shows how each would be coded in the dataset.

Phonological length and lack of acoustic prominence

This section introduces and describes the following predictors:

coding ID: UNSTRESSED_LONG
coding ID: UNSTRESSED_SHORT

In Swedish, the lack of prominence strips a word of its lexical pitch accent. For example, on line 6 of Table 6.2, the 3 in EE3 ([^{2}eː]) is the SweFA code for stress accent 2. The 4 in AH4 ([a$_{2}$]) is the SweFA code for the post-tonic accent following stress accent 2. However, because they are both acoustically non-prominent in this occurrence, their lexical accents are stripped, and they become [$^{2\to0}$eː] and [a$_{2\to0}$], respectively.

If one of the two vowels in the nPVIV pair lacks acoustic prominence and is phonologically long, then the predictor UNSTRESSED_LONG is coded YES as is the case for AA0 [ɑː$_{0}$] on lines 3 and 4, EE3 [$^{2\to0}$eː] on line 6, and EE2 [$^{2\to0}$eː] on lines 40 274 and 40 275. If one of the vowels in the nPVIV pair lacks acoustic prominence and is phonologically short, then the predictor UNSTRESSED_SHORT is coded YES. This is the case for all of the vowel pairs in the snapshot of the dataset (lines 1–9 and lines 40 269–40 277). Table 6.4 provides a list of every possible vowel-pair combination and shows how each would be coded in the dataset.

Phrase-final syllable

This section introduces and describes the following predictor:

coding ID: PHRASE_FINAL

According to Klatt (1976), English vowels lengthen within a pre-pausal foot. This is why Thomas and Carter (2006) and Sarmah et al. (2011) chose to omit phrase-final vowels from their calculation of nPVIV. On the other hand, Low et al. (2000), White and Mattys (2007), Torgersen and Szakay (2012), and Fuchs (2013) chose to include them. White and Mattys (2007) make a strong case for inclusion: firstly, lengthening does not occur before all pauses and secondly, if there is lengthening, it may contribute to the overall perception of rhythm (2007:507).

A number of recent studies have found that phrase-final lengthening may actually be responsible for a significant portion of the rhythmic difference between languages. White et al. (2012) refer to this as the *final lengthening hypothesis*: "utterances can be categorized on the basis of localized lengthening effects, specifically, the degree to which final syllables

			coding in dataset					
			ACCENT_1_LONG	ACCENT_1_SHORT	ACCENT_2_LONG	ACCENT_2_SHORT	UNSTRESSED_LONG	UNSTRESSED_SHORT
accent 1 long	followed by	accent 1 long	YES	NO	NO	NO	NO	NO
	followed by	accent 1 short	YES	YES	NO	NO	NO	NO
	followed by	accent 2 long	YES	NO	YES	NO	NO	NO
	followed by	accent 2 short	YES	NO	NO	YES	NO	NO
	followed by	unstressed long	YES	NO	NO	NO	YES	NO
	followed by	unstressed short	YES	NO	NO	NO	NO	YES
accent 1 short	followed by	accent 1 long	YES	YES	NO	NO	NO	NO
	followed by	accent 1 short	NO	YES	NO	NO	NO	NO
	followed by	accent 2 long	NO	YES	YES	NO	NO	NO
	followed by	accent 2 short	NO	YES	NO	YES	NO	NO
	followed by	unstressed long	NO	YES	NO	NO	YES	NO
	followed by	unstressed short	NO	YES	NO	NO	NO	YES
accent 2 long	followed by	accent 1 long	YES	NO	YES	NO	NO	NO
	followed by	accent 1 short	NO	YES	YES	NO	NO	NO
	followed by	accent 2 long	NO	NO	YES	NO	NO	NO
	followed by	accent 2 short	NO	NO	YES	YES	NO	NO
	followed by	unstressed long	NO	NO	YES	NO	YES	NO
	followed by	unstressed short	NO	NO	YES	NO	NO	YES
accent 2 short	followed by	accent 1 long	YES	NO	NO	YES	NO	NO
	followed by	accent 1 short	NO	YES	NO	YES	NO	NO
	followed by	accent 2 long	NO	NO	YES	YES	NO	NO
	followed by	accent 2 short	NO	NO	NO	YES	NO	NO
	followed by	unstressed long	NO	NO	NO	YES	YES	NO
	followed by	unstressed short	NO	NO	NO	YES	NO	YES
unstressed long	followed by	accent 1 long	YES	NO	NO	NO	YES	NO
	followed by	accent 1 short	NO	YES	NO	NO	YES	NO
	followed by	accent 2 long	NO	NO	YES	NO	YES	NO
	followed by	accent 2 short	NO	NO	NO	YES	YES	NO
	followed by	unstressed long	NO	NO	NO	NO	YES	NO
	followed by	unstressed short	NO	NO	NO	NO	YES	YES
unstressed short	followed by	accent 1 long	YES	NO	NO	NO	NO	YES
	followed by	accent 1 short	NO	YES	NO	NO	NO	YES
	followed by	accent 2 long	NO	NO	YES	NO	NO	YES
	followed by	accent 2 short	NO	NO	NO	YES	NO	YES
	followed by	unstressed long	NO	NO	NO	NO	YES	YES
	followed by	unstressed short	NO	NO	NO	NO	NO	YES

Table 6.4: Exhaustive inventory of all possible pitch-accent and phonological-quantity combinations upon which an nPVIV calculation can be based. The coding for each pair is provided by means of the six columns to the right.

are longer than those in utterance-medial position" (2012:676). The authors created a series of *sasasa* guises whereby all consonants were substituted with /s/, all vowels with /a/, fundamental frequency neutralized, and speech rate normalized. They then created stimuli that included the final foot and stimuli with the final foot removed. Listeners relied on phrase-final lengthening to distinguish between the varieties that actually had that feature: Standard Southern British English and Welsh English. For the languages without substantial lengthening – Castilian Spanish and Orkney Islands English – listeners made no distinction in assessing stimuli with and without the final foot (2012:676). In other words, speakers of languages with phrase-final lengthening adapt to its presence and use it as a 'shortcut cue'. Prieto, del Mar Vanrell, Astruc, Payne, and Post (2012) also found that phrase-final lengthening was more extensive in English than in Spanish or Catalan. In their view, it explained, in part, the rhythmic differences between the former and the latter (2012:694).

Like English, Swedish has phrase-final lengthening, but it is unclear what its relationship is to perceptions of rhythm. But the connection between phrase-final lengthening and rhythm has been sufficiently established in other literature to motivate its place in the statistical model. Recall also that Bodén (2010:72–75) found that speakers of multiethnolect in both Malmö and Gothenburg engage in more phrase-final lengthening than speakers of the respective regional mainstream varieties (refer to p. 104).

Therefore, I code those vowels into the dataset and include them as predictors in the model. The column ID headed PHRASE_FINAL is coded YES if it occurs within the syllable before a pause. A pause is defined as any break of 150 milliseconds or more. Lines 5 and 40 273 in Table 6.2 exemplify two phrase-final vowels.

Syllable-final /r/

This section introduces and describes the following predictor:

coding ID: CODA_R

Recall on p. 130 that in Stockholm Swedish, syllable-final rhotics will merge with the following consonant to produce a retroflex allophone in the cases where the following consonant is /d, n, s, t/. This sandhi process renders [ɖ, ɳ, ʂ, ʈ], respectively. Where the following syllable begins with a vowel, the rhotic becomes the onset of that syllable:

förhoppningar om ⟶ förhoppninga rom (Eng.: hopes about)

For the remaining scenarios, two forms are possible: (1) the rhotic is either elided or (2) is expressed – usually as an approximant:

(1) irakier på flykt ⟶ irakie på flykt (Eng.: Iraqis fleeing)

(2) irakier på flykt ⟶ irakieɹ på flykt

Because the boundary between a vowel + approximant is quite subjective, I combine their duration and treat them as a single segment. Below is an example in SweFA notation:

irakier på flykt ⟶ IH0. R AA1. K IH0. EH0_R. P OA1. F L YH1 K T

This, however, renders extremely long segments, which can render high contrasts, which can skew the calculation of nPVIV if certain speakers happen to have more syllable-final rhotics than others. Therefore, these occurrences are coded YES in the model under the heading CODA_R. Line 1 in Table 6.2 illustrates an example.

Speech rate

This section introduces and describes the following predictor:

coding ID: SPEECH_RATE

Speech rate has been shown to affect rhythm even in cases where rhythm is operationalized by means of a normalizing algorithm. The literature is consistent in showing that higher rates of speech result in generally less durational contrast (Dellwo 2010; Deterding 2001; Torgersen & Szakay 2012; although cf. Fuchs 2013). In the case of London's multiethnolect, Torgersen and Szakay (2012:831–832) have demonstrated a correlation between higher rates of speech and lower nPVIV. Arvaniti (2009) has suggested that it might be the spacing of prominences in rapid speech may give it that 'syllable-timed' impression:

> The combination of relatively stable interstress intervals and different speaking rates suggests that languages like Greek, Spanish or Italian include more syllables in each interval, an observation confirmed by Dauer (1983). Thus, one difference between languages called stress-timed and those called syllable-timed may have to do with the spacing of prominences, not in terms of duration but in terms of number syllables; in this respect, prominences may be sparser in syllable-timed languages. In turn, this difference could result in different rhythmic hierarchies and different degrees of flexibility in keeping prominences regular. (Arvaniti 2009:59)

I think a more likely explanation is that there is a margin of diminishing returns on contrast whereby faster speech hits a limit on vowel reduction. Particularly with long-short combinations like /ˈmɑːˌta/ ('mata', *feed*), the reduction of /a/ may hit a limit at higher speeds, forcing /ɑː/ to be further reduced, thereby changing their durational ratio. At lower speeds, both phonemes may simply reduce by equal proportions. This sort of trend is reflected in Thomas and Carter (2006:348), reproduced in Figure 6.2. There, the trend appears to be logarithmic; e.g., small accelerations in rapid speech will reduce nPVIV more than small accelerations in slow speech, and small decelerations in rapid speech will increase nPVIV more than small decelerations in slow speech. The same trend appears in this dataset. Figure 6.3

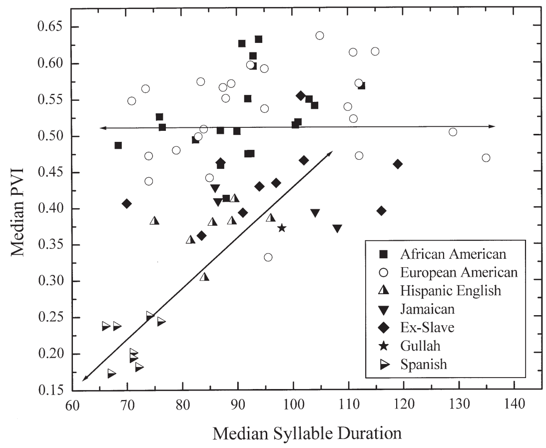


Figure 6.2: Reproduction of Thomas and Carter (2006:348): The relationship between speech rate and nPVIV for varieties of English and Spanish

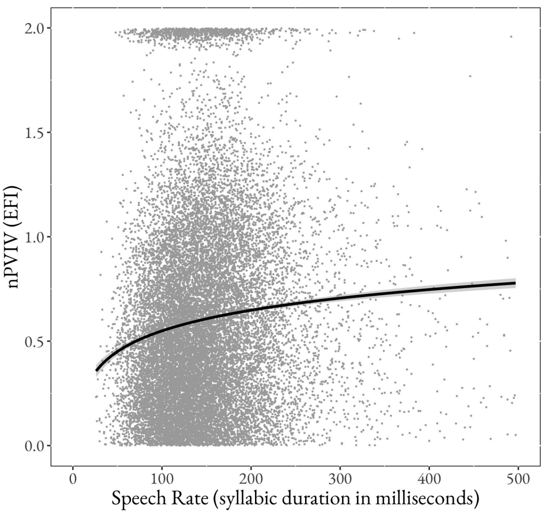


Figure 6.3: The relationship between speech rate and nPVIV within the CASUAL-speech data subset (n = 21 000).

shows a scatterplot of the influence of rate on nPVIV for the 21 000 CASUAL-speech observations. The relationship between rate and nPVIV is nonlinear and appears to be strongest at higher speech rates where $x < 120$. Therefore, I calculate its natural logarithm for the statistical model.

I measure rate as the mean duration of the two syllables that contain each vowel pair. This takes from Thomas and Carter (2006) and Deterding (2001) who also represent speech rate with syllabic duration. It differs, however, from Torgersen and Szakay (2012:828) who calculate rate as the "number of vocalic elements divided by the duration of all vocalic elements in the particular recording" and Dellwo (2010:5) who calculates rate as the number of segments per second[6].

[6]It is my view that how one measures rate is less important when analyzing a single language. Syllable duration

Lexical frequency

This section introduces and describes the following predictor:

coding ID: LEXICAL_FREQ

Lexical frequency can affect vowel reduction (Pluymaekers, Ernestus, & Baayen 2005) as well as fuller realizations of segments (Zhao & Jurafsky 2009), both of which can affect rhythm. For example, in my interview with participant Shorty, he used a high number of medical terms in a story about losing a leg in a drive-by shooting. He also had a higher nPVIV than one would expect in spontaneous multiethnolectal speech. Similarly, participant Paul – another high-nPVIV speaker – recounts his experience of seeing 'gays in gas masks' in Gran Canaria. His nPVIV was also quite high. The possibility that low-frequency words could minimize reduction and increase rhythmic contrast suggests that lexical frequency needs to be accounted for in any statistical model of rhythm.

Naturally, the READING styles (analyzed in the next Chapter) control for lexical frequency because the participants all read the same passage. In the CASUAL dataset, however, the lexical inventory varies. I obtained frequency data from the *Swedish Spoken Language Corpus* (SSLC) at Gothenburg University (Allwood 1999) and coded each word in my dataset accordingly. Despite being a modest corpus of 1 230 663 words, the SSLC is the largest modern spoken corpus of Swedish that I am aware of.

Lexical frequency was coded for each vowel pair in the column headed LEXICAL_FREQ (Table 6.2). Words that occurred in my data that were missing from the SSLC were given a frequency of 1, because the speaker must have heard the word at least once in order to produce it. Vowel pairs that straddle two words take the mean frequency of those two words. For example, on line 40 277 in Figure 6.2, the word frequency for EH0_YH0 comes from *välfungerande system* ('well-functioning system'). *Välfungerande* is missing from the SSLC[7], so I coded it as 1. *System* occurs in the SSLC 81 times. The mean of 1 and 81 is 41.

Since lexical frequency is Zipf-distributed, its natural logarithm is used in the statistical model (Pluymaekers et al. 2005; Zhao & Jurafsky 2009).

Random effects

This section introduces and describes the following random effects:

(1|SPEAKER)
(1|VOWEL)

or syllables per second or segmental duration or segments per second all seem to be adequate approaches. The choice of measurement gains importance only once one cross-compares languages, because they can have different syllable structures (consider languages that have CV syllables versus CCCVCCC) (Dellwo 2010:67).

[7] *Väl* and *fungerande*, however are in the SSLC. One shortcoming to my coding is that it did not take into account the components of less-frequent compound words. The corpus does, however, include common compound words.

	Internal factors (CASUAL) - Summary												Speech rhythm (CASUAL)	
Group / social class	Pseudonym	Profession	*n* accent-1 long vowels	*n* accent-1 short vowels	*n* accent-2 long vowels	*n* accent-2 short vowels	*n* unstressed long vowels	*n* unstressed short vowels	*n* phrase-final vowels	*n* /r/ in coda position	mean speech rate in milliseconds	mean lexical frequency	*n* vocalic intervals	mean nPVIV (CASUAL)
'invandrare', working	Reman	social worker	55	37	12	29	159	254	26	29	150	7 641	340	55.8
	Malik	cook at cafeteria, street food	77	55	18	34	117	253	48	7	147	7 019	333	48.0
	Max	youth worker	149	86	44	71	318	565	57	62	134	7 004	742	48.8
	Antonio	caretaker, ex drug dealer	166	96	73	124	307	659	63	65	160	5 927	849	45.6
	Hayder	caretaker, bouncer	85	59	38	77	224	403	36	24	160	6 243	539	54.8
	Dawit	group home assistant	195	82	73	93	361	761	78	56	141	5 583	961	51.3
	Solomon	youth worker	150	120	60	112	370	820	61	74	137	7 776	993	59.0
	Murad	group manager, furniture retail	98	64	36	58	180	389	43	14	137	5 190	504	45.2
	Abel	photographer, personal trainer	110	53	30	61	201	383	52	34	149	7 800	500	54.6
	Mateo	waiter, bartender	93	63	33	48	278	432	60	43	145	7 502	585	60.6
	Reza	waiter, napropathy student	120	55	46	39	230	457	39	28	154	5 634	583	51.9
	Mezdar	youth worker (certified)	118	45	21	51	185	369	47	46	150	6 893	473	52.4
	Jocke	athletic coach, personal trainer	194	100	62	100	450	826	90	70	144	7 604	1 065	57.0
	Sohrab	web architect	149	67	37	61	298	548	67	48	151	6 529	714	63.5
	Shorty	social worker	89	82	22	45	195	325	62	30	137	8 545	455	52.1
	Kevin	organisational consultant	137	67	40	64	316	597	54	42	140	7 532	761	58.1
	Tarik	field assistant; athletic coach	108	89	37	56	200	454	51	39	151	5 168	573	49.2
	Majeed	doctor	133	85	32	55	278	544	59	36	134	7 952	693	65.4
'invandrare', upper	Parviz	business owner, physician	116	40	33	54	253	349	45	34	148	8 989	507	57.5
'svensk', working	Per	demolition operative	66	28	27	36	205	363	43	27	153	9 555	453	78.3
	Rasmus	chef at major hotel	74	48	39	51	213	361	57	25	141	7 702	478	72.3
	Johnny	waiter, bartender	115	88	51	58	251	443	63	47	166	6 541	591	64.6
	Sven	team leader, security firm	50	38	29	36	135	324	29	38	151	8 703	379	61.5
	Paul	self-employed, renovations	91	67	32	46	240	414	62	54	156	7 528	543	74.3
	Jesper	waiter	61	49	35	64	225	427	45	38	132	7 672	541	64.8
	Richie	pest exterminator	112	77	50	44	284	486	37	57	149	9 353	637	60.4
	Thomas	middle office associate, hedgefund	173	68	35	65	376	648	54	62	140	8 233	822	65.6
	Nils	supermarket clerk	100	62	23	51	208	456	44	34	152	8 233	565	57.2
	August	section manager, IT	66	58	29	33	251	352	30	39	149	8 025	482	67.0
	Jonte	middle school teacher, special ed	113	41	55	60	263	513	47	34	144	6 249	640	71.2
	Martin	massage therapist, former teacher	101	37	38	31	287	422	41	23	148	8 346	557	91.0
	Gunnar	high school teacher	118	51	30	43	235	390	52	38	144	7 395	517	60.0
	Joseph	communications strategist/PR	84	47	24	37	148	296	38	23	178	7 675	378	56.0
	Jan-Bertel	portfolio manager, hedgefund	126	65	39	44	292	433	69	28	132	6 478	584	61.6
	Johan	chief marketing officer	38	17	29	34	82	173	20	22	162	9 098	220	57.5
'svensk', upper	Jan-Axel	trader, hedgefund	92	48	45	45	154	347	61	35	170	7 633	443	51.7
		Means:	*109*	*62*	*38*	*56*	*244*	*451*	*51*	*39*	*148*	*7 415*	*21 000*	57.8
		Percentages:	19%	11%	6%	10%	42%	77%	9%	7%				

Table 6.5: Overview of dataset for internal predictors: pseudonym, profession, total counts for accent 1 and 2 long and short vowels, unstressed long and short vowels, vowels in phrase-final position, rhotics in coda position, mean speech rate in milliseconds, mean lexical frequency, the number of vocalic intervals counted for nPVIV, and the mean nPVIV. Speakers are arranged by the racialized *invandrare*/*svensk* binary and social class (discussed in Sections 7.1 and 7.2).

Speaker For this dataset, speaker is a particularly important random effect because the number of observations per speaker varies by a considerable amount. The lowest observation count is 295 for August (pseudonym) in casual speech, and the highest observation count is 1517 for Jan in casual speech. In a simple linear-regression model, the social factors associated with Jan would weigh five times more than those associated with August. In a mixed-effects model, including the random effect of speaker will mathematically adjust for this bias.

Vowel Every observation-level nPVIV consists of a vowel pair. These pairs are not unique for every observation; rather, there are 3 332 unique pairs within the 40 277 nPVIV calculations. Vowel combination is a particularly important random effect to capture in this dataset because the reading passage (to be discussed in Chapter 8) only has 632 unique vocalic combinations compared with the 3 099 unique vocalic combinations in casual. Two hundred thirty-three of the vocalic pairs found in the reading passage – 33% of its total pairs – are not found at all in the casual-speech corpus. Therefore, there would be a danger that contextual style effects would manifest themselves as vowel effects if they are not accounted for in the statistical model.

Data summary of internal predictors

Table 6.5 provides a summary of the casual-speech dataset analyzed here. Although the focus in this section is internal factors, speakers are arranged by social class and racialization since those are the predictors that the main analysis is based on (discussed in Sections 7.1 and 7.2). It includes pseudonyms, profession, aggregated internal factors for each speech style, the number of vocalic pairs calculated, and the mean nPVIV. Note, however, that these are participant means and that the actual models are run on the full dataset, which is too large to provide in a figure. As Thomas and Carter (2006) note, means are not always the best way to represent the central tendencies of skewed measurements like nPVIV (2006:343). Therefore, I have provided a copy of the same table with medians (including data from the other styles) on page 220.

Statistical model

I built a mixed-effects linear regression model, using *lme4* in *R*, with the following call:

> nPVIV ~ accent_1_long + accent_1_short + accent_2_long + accent_2_short + unstressed_long + unstressed_short + pair_coda_r + phrase_final + *log*(lexical_freq) + *log*(speech_rate) + (1|speaker) + (1|vowel).

Variance inflation factor (VIF) was calculated separately using the *car* package. Table 6.6 contains the model output.

The model in Table 6.6 predicts that all of the internal predictors have an effect on nPVIV. The first eight predictors are binary, and the model's reference point is their presence, indi-

	Model 1 *Internal only*	
Response variable	nPVIV (casual)	
(Intercept)	7.8 (4.7)$^{\circ}$	
Internal predictors		VIF
accent_1_long·*yes*	23.6 (1.0)***	1.4
accent_1_short·*yes*	14.6 (1.2)***	1.2
accent_2_long·*yes*	18.3 (1.5)***	1.2
accent_2_short·*yes*	10.7 (1.2)***	1.2
unstressed_long·*yes*	8.6 (1.0)***	1.8
unstressed_short·*yes*	15.5 (1.1)***	1.7
coda_r·*yes*	7.9 (1.3)***	1.0
phrase_final·*yes*	6.9 (1.1)***	1.0
log(lexical_freq)	−0.3 (0.1)**	1.1
log(speech_rate)	6.0 (0.8)***	1.1
AIC	218 859.3	
BIC	218 970.6	
Log Likelihood	−109 415.6	
Num. obs.	21 000	
Num. groups: vowel	3 099	
Num. groups: speaker	36	
Var: vowel (Intercept)	70.1	
Var: speaker (Intercept)	96.3	
Var: Residual	1 906.1	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Table 6.6: Mixed-effects linear regression model with nPVIV in casual speech as the response variable and internal factors as the predictors. For categorical predictors, the reference category is in italics. Coefficients are indicated in the center column, standard errors in the parentheses, and variance inflation factors (VIF) to the right.

Estimated nPVIV for a typical vowel pair within a typical word at a typical speech rate:

= [*base constant*] + [*accent-length coefficient*]

= *Intercept* + *coefficient*(*lexical_freq*) + *coefficient*(*speech_rate*)+ [*accent-length coefficient*]

= [7.8] + [−0.3 · *ln*(7000)] + [6.0 · *ln*(150)]+ [*accent-length coefficient*]

= [7.8] + [−2.7] + [30.1]+ [*accent-length coefficient*]

= 35.2+ [*accent-length coefficient*]

Table 6.7: Calculation of a *base constant* for a 'typical' vowel pair, as predicted by the model in Table 6.6. The assumption is that the word frequency is 7000, speech rate is 150 milliseconds, and that phrase-finality and an /r/ in coda position are not present. This *base constant* is added to the *accent-length coefficient* in Table 6.8 in order to demonstrate the internal effects that accent (or the lack thereof) has on nPVIV in the dataset.

Stress-length combinations within pairs		Accent-length coefficient		Estimated nPVIV			*n* in dataset	*% n* in dataset	
→accent-1 long V + unstressed short V	23.6 + 15.5 =	39.1	+35.2 =	74.3	*high*	*less*	2 472	11.8	←
accent-2 long V + unstressed short V	18.3 + 15.5 =	33.8	+35.2 =	69.0	↑	↑	972	4.6	
accent-1 long V + unstressed long V	23.6 + 8.6 =	32.4	+35.2 =	67.4			1 211	5.8	
accent-1 short V + unstressed short V	14.6 + 15.5 =	30.1	+35.2 =	65.3			1 326	6.3	
accent-2 long V + unstressed long V	18.3 + 8.6 =	26.9	+35.2 =	62.1			268	1.3	
accent-2 short V + unstressed short V	10.7 + 15.5 =	26.2	+35.2 =	61.4	alternation	staccato	1 286	6.1	
→unstressed short V + unstressed long V	15.5 + 8.6 =	24.1	+35.2 =	59.3			4 576	21.8	←
accent-1 short V + unstressed long V	14.6 + 8.6 =	23.2	+35.2 =	58.4			687	3.3	
accent-2 short V + unstressed long V	10.7 + 8.6 =	19.3	+35.2 =	54.5			485	2.3	
→unstressed short V + unstressed short V	15.5	15.5	+35.2 =	50.7	↓	↓	5 604	26.7	←
unstressed long V + unstressed long V	8.6 =	8.6	+35.2 =	43.8			1 542	7.3	
remaining combinations		*n/a*		*n/a*	*low*	*more*	571	2.7	
						total:	21 000	100	

Table 6.8: Vowel-pair combinations in dataset rank-ordered by effect on nPVIV (highest to lowest), as predicted by model in Table 6.6 above. Estimated nPVIV is calculated using a word frequency of 7000 and a speech rate of 150 milliseconds. Vowel-pair combinations that constitute > 10% of the total observations are marked with an arrow.

cated by *yes*. The remaining two parameters are both nominal and logarithmic. Table 6.7 calculates a case study *base constant* for the rhythm of a vowel pair based on a word frequency of 7 000, and an average speech rate of 150 milliseconds. From this base constant, estimated nPVIV values are calculated in Table 6.8.

Table 6.8 aims to show two things. First, there is a high degree of internally-governed variation regarding rhythmic contrast. The accent-length combinations are listed in rank order according to their effects on nPVIV. The highest nPVIV values (i.e.. *high alternation* and *not staccato*) are predicted to occur when accented vowels are adjacent to unstressed vowels. Long vowels taking accent 1 adjacent to unstressed short vowels are predicted to render the highest alternation, with an estimated nPVIV of 74.3. Refer to Table 6.4 and recall that an accent-1 vowel adjacent to an unstressed short vowel is coded as follows.

			ACCENT_1_LONG	ACCENT_1_SHORT	ACCENT_2_LONG	ACCENT_2_SHORT	UNSTRESSED_LONG	UNSTRESSED_SHORT
accent 1	followed by	unstressed short	YES	NO	NO	NO	NO	YES
unstressed short	followed by	accent 1	YES	NO	NO	NO	NO	YES

On the other hand, the lowest nPVIV values (i.e., *low alternation* and *staccato*) are predicted to appear when unstressed vowels are adjacent to other unstressed vowels. Unstressed long vowels adjacent to unstressed long vowels are predicted to render the lowest rhythmic alternation, with an estimated nPVIV of 43.8. Refer to Table 6.4 and recall that an unstressed long vowel adjacent to an unstressed long vowel is coded as follows.

			ACCENT_1_LONG	ACCENT_1_SHORT	ACCENT_2_LONG	ACCENT_2_SHORT	UNSTRESSED_LONG	UNSTRESSED_SHORT
unstressed long	followed by	unstressed long	NO	NO	NO	NO	YES	NO

In other words, the internally-constrained envelope of rhythmic variation is quite large and lies between approximately 43.8 on the lower end and 74.3 on the upper end.

Second, only three types of combinations are actually frequent. These combinations constitute 60.3% of all total occurrences and are arrowed in Table 6.8. Long vowels taking accent 1 adjacent to unstressed short vowels occur in 11.8%, unstressed short vowels adjacent to unstressed long vowels occur in 21.8%, and unstressed short vowels adjacent to other unstressed short vowels occur in 26.7% of total instances. This is important to note for the following reason. If I find any socially-governed variation for overall nPVIV, I will also need to examine how social predictors influence these three categories in isolation. That analysis will occur in Chapter 9.

2.3 Section summary

This section has presented two key observations. First, it offers a statistical model of rhythm that takes into account the main (internal) factors that contribute to rhythmic alternation in Swedish prosody. In subsequent sections, this model will be used as a base for testing the influence of social factors. Second, it shows how much variation exists absent of any social factors. Rhythmic variation appears to be primarily governed by the distribution of prominence and phonological quantity. I constructed Table 6.8 in order to demonstrate how systematic it is. In general, prominent (accented) vowels adjacent to non-prominent (unstressed) vowels show the highest nPVIV values (i.e., *less staccato*). This is magnified if the prominent vowel is phonologically long and the non-prominent vowel is phonologically short. On the other end of the spectrum, non-prominent vowels adjacent to non-prominent vowels show the lowest nPVIV values (i.e., *more staccato*). Absent of speaker effects and social factors, the model predicts an nPVIV range that is substantial – between 43.8 and 74.3.

So how is it that rhythm can really even be a sociolinguistic variable? If rhythm has, in fact, been coerced for social meaning, it will be necessary to identify where, specifically, the coercion is occurring. This is something I will return to at a later point in the dissertation (Chapter 9). The goal of this introductory analysis was to identify concrete components behind what is otherwise an intangible construction. In the following section, I will expand the model that I ran in Table 6.6 by adding social parameters. This will be the first step in shedding light on whether rhythm, is in fact, socially differentiated in Stockholm.

3 Introducing classic social predictors to the model

I return to the research question: *Does the speech of Stockholm's racialized working class have lower rhythmic alternation than other varieties in the city; i.e., is it 'staccato'?*. By asking this question, I hope to shed light on what sometimes is referred to as Swedish multiethnolect, a variety ascribed to young, working-class men from Stockholm's multiethnic periphery. This definition is most explicitly linked to the parameters age, racialization, and social class. Age operationalizes in a straightforward manner. Social class and racialization, however, are challenging constructs. The theoretical framework that I use for each was therefore laid out in detail in Methods Chapter 4 on pages 86–95.

Following this analysis of age, racialization, and social class, Section 4 will examine the social-network data from the 29 participants who chose to do the second interview.

3.1 Operationalizing age and racialization

This section introduces and describes the following predictor:

coding ID: AGE
coding ID: RACIALIZATION

I used the participant's age in 2017. This was preferable over actual birth year, because the other nominal predictors in the model are normalized to a 100-point scale whereas birth year lies on a 2000-point scale.

Racialization is coded into the statistical model as one of two categorical variables: *invandrare* or *svensk*. Refer to pages 86–89 for a thorough explanation of each.

3.2 Operationalizing social class

The following subsections outline how each social predictor is operationalized: *occupational status*, *income*, *formal education*, *parental occupational status*, *parental formal education*, and *taste*.

Occupational status

This section introduces and describes the following predictor:

coding ID: SEI

I measure occupational status with Ganzeboom and Treiman's (2003; 2018) International Socio-Economic Index of occupational status (ISEI). It is a 100-point numerical status rank of the 540 most common Western professions.

The list of professions in the index comes from the 2008 International Standard Classification of Occupations (ISCO-08), which was established by the Governing Body of the International Labour Office in accordance with the Seventeenth International Conference of Labour Statisticians (ICLS) (International Labour Office 2012). The 100-point status index is based on the most recent International Stratification and Mobility File (ISMF), a survey developed by Ganzeboom and Treiman (2003). The ISEI index is derived from the survey subjects' education, income, and wealth along with his/her father's education and income attainment. It was calculated in 1968, 1988, and 2008 from extensive demographic information that linked the aforementioned variables to specific ISCO occupations. The procedure for this was last refined for the 1988 index, which means that the 2008 index was calculated using the same methods as in 1988, albeit with the new 2008 data. Occupational status is considered independent of the other factors that make up an individual's class standing:

> Our preferred way to think about SEI is that it measures the attributes of occupations that convert a person's main resource (education) into a person's main reward (income). A simple model of the stratification process looks like this:
>
> EDUCATION ⟶ OCCUPATION ⟶ INCOME
>
> Occupation can be regarded as an intermediate position – similar to a latent variable – that converts education into income. In this interpretation, SEI is not so much a consequence of true occupational status as an approximation to it. (Ganzeboom et al. 1992:8–9)

	Hout, Smith & Marsden (2015)				Ganzeboom & Treiman (2003)	
	PRESTG105	PRESTG105PLUS	PRESTG10	SEI10	SIOPS	ISEI
Ulfsdotter Eriksson (2006)	*0.77*	*0.77*	*0.78*	*0.81*	*0.84*	*0.85*

Table 6.9: Pearson correlation coefficient of PRESTG105, PRESTG105PLUS. PRESTG10, SEI10, SIOPS, and ISEI with Ulfsdotter Eriksson's (2006) prestige list of 100 Swedish professions. ISEI rendered the highest correlation and was therefore selected for this study.

As far as I am aware, no comprehensive index exists for Sweden[8]. Carlsson (1958:146–149) produced a status hierarchy for 26 professions based on the survey responses of 1 700 participants. More recently, Ulfsdotter Eriksson (2006) ranked 100 common Swedish professions by prestige according to the survey responses of over 3000 participants. However, neither of the aforementioned studies cover enough professions to be useful for my study.

Several non-Swedish global indices circulate that are built from this decade's demographic data (2010–2020). Hout, Smith, and Marsden (2015) used the methods established in Hauser and Warren (1997) to calculate a socioeconomic index for the recently-released 2010 United States census data (SEI10). They also developed three prestige indices of their own (PRESTG10, PRESTG105, PRESTG105PLUS). Ganzeboom and Treiman (2003, 2018) developed two socioeconomic indices, The Standard International Occupational Prestige (SIOPS) and The International Socio-Economic Index of Occupational Status (ISEI).

I selected ISEI because it correlated most closely to Ulfsdotter Eriksson's (2006) prestige ranking that I cite above. I conducted six separate Pearson correlations for each of the six indices against Ulfsdotter Eriksson's (2006) Swedish prestige ranking. As shown in Table 6.9, ISEI correlates the most highly.

I coded each participant with the ISEI that corresponded to their self-reported profession. If the participant worked several part-time jobs, I calculated an average ISEI for those jobs.

Income

This section introduces and describes the following predictor:

coding ID: INCOME

Participants were asked to self-report their monthly income category before tax. In the statistical model, income was coded as follows:

[8]Ulfsdotter Eriksson (2006) makes the same observation in her literature review: no index is in circulation (2006:34).

Swedish kronor	Euro	Coding
0 - 10 000 SEK	€0 - €1 000	5
10 000 - 20 000 SEK	€1 000 - €2 000	15
20 000 - 30 000 SEK	€2 000 - €3 000	25
30 000 - 40 000 SEK	€3 000 - €4 000	35
40 000 - 50 000 SEK	€4 000 - €5 000	45
50 000 - 60 000 SEK	€5 000 - €6 000	55
60 000 - 70 000 SEK	€6 000 - €7 000	65
70 000 - 80 000 SEK	€7 000 - €8 000	75
80 000 - 90 000 SEK	€8 000 - €9 000	85
90 000 - 100 000 SEK	€9 000 - €10 000	95
more than 100 000 SEK	*more than* €10 000	100

Formal education

This section introduces and describes the following predictor:

coding ID: EDUCATION

In the statistical model, education was coded as follows:

Formal Education	Coding
less than primary school	1
primary school	16.7
secondary school	33.3
post-secondary vocational training	50
1 to 3 years of university	66.7
4 to 5 years of university	88.3
Ph.D.	100

Parental occupational status

This section introduces and describes the following predictor:

coding ID: PARENTAL_SEI

Participants provided me with the last-known occupation of their parent or parents. Occupation was coded numerically according to the ISEI as described on page 145. If the parent worked several part-time jobs, I calculated an average ISEI for those jobs. If the participant was raised by two parents, I used the highest SEI of the two parents.

Parental formal education

This section introduces and describes the following predictor:

coding ID: PARENTAL_EDUCATION

Group	Description	%	%	Typ	Description	%	%	
A	Affluent Pioneers	5.24	4.00	A01	Scandinavia's Wealthiest	0.70	0.52	
				A02	Dream Weavers	1.12	0.83	
				A03	Serious Money	2.25	1.64	
				A04	Careers and Kids	1.18	1.00	highbrow
B	Metropolitan Pioneers	5.94	6.29	B05	Cosmopolitan Families	0.64	0.53	
				B06	Ethical City Families	0.59	0.50	
				B07	Mid-Career Climbers	1.46	1.44	
				B08	Culture and Politics	1.84	2.05	
				B09	Uptown Singles	1.40	1.77	
C	Young Urban Pioneers	6.48	7.13	C10	Young Cosmopolitans	2.01	2.44	
				C11	Urban Optimists	2.39	1.95	
				C12	First Foot on the Ladder	2.07	2.75	
D	Curious Pioneers	5.92	6.95	D13	Side Street Singles	1.95	2.48	
				D14	Unattached Multi-Cultures	2.85	3.02	
				D15	University Fringe	0.62	0.72	
				D16	Study Buddies	0.50	0.73	
E	Family Oriented Followers	9.32	7.59	E17	Urban Commuter Families	4.21	3.46	
				E18	Solid Suburban Life	5.10	4.14	
F	Struggling Followers	7.57	8.67	F19	Bright Young Things	2.06	2.16	lowbrow
				F20	Multicultural Towers	2.45	2.47	
				F21	Ethnic Urban Life	3.07	4.04	
G	Multicultural Followers	6.50	5.91	G22	Lower-income Essentials	3.21	3.08	
				G23	Multicultural Families	3.29	2.83	

Figure 6.4: *Mosaic Sweden*'s 23 urban market segments for Sweden (duplication from Experian Ltd and InsightOne Nordic AB 2013:3). Five market segments are represented in my dataset: *A01 Scandinavia's Wealthiest*, *B06 Ethical City Families*, *B08 Culture and Politics*, *D15 University Fringe*, and *G23 Multicultural Families*. Each of the five segments are marked with an arrow. The top three segments are coded as *highbrow*; the bottom two as *lowbrow*.

Participants provided me with the highest-level of formal education for their parent or parents. Formal education was coded numerically according to the same schema described in the above section. If the participant was raised by two parents, I calculated an average for the two parents.

Taste

This section introduces and describes the following predictor:

coding ID: MOSAIC_TASTE

Consumer market segmentation is a taxonomy that identifies taste-based population subgroups for the purpose of supplying targeted goods and services to them. In this respect, market segments can be seen as the statistical clustering of Bourdieu's (1984) *habitus*, the term he uses for "the systems of dispositions characteristic of the different classes and class fractions" (1984:6).

> Taste classifies, and it classifies the classifier. Social subjects, classified by their classifications, distinguish themselves by the distinctions they make, between the beautiful and the ugly, the distinguished and the vulgar, in which their position in the objective classifications is expressed or betrayed. And statistical analysis does indeed show that oppositions similar in structure to those found in cultural practices also appear in eating habits. The antithesis between quantity and quality, substance and form, corresponds to the opposition — linked

> to different distances from necessity — between the taste of necessity, which favours the most 'filling' and most economical foods, and the taste of liberty — or luxury — which shifts the emphasis to the manner (of presenting, serving, eating etc.) and tends to use stylized forms to deny function. (Bourdieu 1984:6)

Adli (2013) found strong correlations between lifestyle choices (such as going to parties or reading books) and syntactic variation in French *Wh*-questions. Whereas he developed his own inventory of lifestyle questions, I chose to use 60 consumer interests developed by Experian Ltd and InsightOne Nordic AB (2013). Experian collects data around the world on consumer behavior and identifies market segments for each country by conducting cluster analyses on those behaviors. Its Swedish partner InsightOne examined the actual behavior (by tracking consumer purchases) of consumers and combined this data with approximately 44 000 responses to the *Orvesto* survey (personal communication with InsightOne, 2016). Based on the clustering of 60 consumer interests, it created 44 unique market segments, entitled *Mosaic Sweden*, 23 of which are predominantly urban (Experian 2013:3). Figure 6.4 shows these 23 urban market segments. The interests upon which *Mosaic Sweden* is built include:

gardening	*astrology*	*diet tips*
visiting sports events	*cinema/audio equipment*	*computers*
union issues	*football*	*outdoor life*
follow popular TV-shows	*associations*	*golf*
exercise classes/gym	*interior decorating*	*health and healthcare*
healthy foods	*ice hockey*	*hunting*
equality	*martial arts*	*art*
royal families	*celebrities*	*literature*
skiing	*reading books*	*crosswords*
sudoku	*cooking*	*meditation/yoga*
repairing cars	*environmental care*	*fashion*
exercise	*motor sport*	*new technology*
national economy	*private economy*	*trying new products*
religion	*horse riding*	*watch sport on TV*
sailing	*holidays in Sweden*	*holidays abroad*
disease and health care	*tax*	*DIY*
computer games	*sport fishing*	*needlework*
theatre	*pony trekking*	*youth culture*
downhill skiing	*foreign policy*	*vocational training*
wine tasting	*motorboats*	*extreme sports*

I created a survey with the above interests and asked the participants to mark YES or NO. I then assigned them a *Mosaic* segment based on the statistically closest match between their interest combination and the combinations provided for each *Mosaic* segment. Appendix D contains the full inventory of their interests along with the market segment each participant was assigned to[9]. Five *Mosaic* segments are represented in my data set: A01 Scandinavia's Wealthiest, B06 Ethical City Families, B08 Culture and Politics, D15 University Fringe, and G23 Multicultural Families. As shown in Figure 6.4, these five segments are bimodally distributed. My interpretation is that the former three constitute highbrow segments, and the latter two constitute lowbrow segments. I used the binary categorical coding *highbrow* and *lowbrow* for this parameter.

Data overview

Table 6.10 provides an overview of the data with summaries for each predictor that will be used in the subsequent analysis. Speakers are grouped by the *svensk*/*invandrare* racialized binary and then ranked by social class (PC1) within each group.

3.3 Analysis

This section will examine the effects of age, racialization, and the components of social class on nPVIV by adding these predictors to the base model built in Section 2.2 (Table 6.6). The components of social class will be examined via two approaches: (1) in aggregate by means of a Principle Components Analysis (PCA) and (2) individually to see whether there is any distinction between more embodied (parental SEI, parental education, Mosaic taste) and less embodied components of class (SEI, income, education).

Evaluating the class components in aggregate

Developed originally for research in educational psychology, Hotelling (1933) describes his Principle Component Analysis (PCA) with the following example:

> Consider n variables attaching to each individual of a population. These statistical variables $x_1, x_2, \ldots, x_n$ might for example be scores made by school children in tests of speech and skill in solving arithmetical problems or in reading [...]. The x's will ordinarily be correlated. It is natural to ask whether some

[9]It is important to note that Experian built its market segments on more than just interests and purchasing behavior. It factored in other class-related data, including income, post code, age, and so on. However, my participant assignment is based *solely on interest*. So for example, participant August's assignment to *A01 Scandinavia's Wealthiest* means his tastes resemble that of the several thousand consumers in that segment more than that of consumers in other segments. It does not necessarily mean, however, that his overall income-residence-taste combination fits *A01 Scandinavia's Wealthiest*. I chose to exclude other factors in my calculation in order to keep the *taste* parameter orthogonal.

			Internal factors (CASUAL) - Summary										Social factors - Summary									Speech rhythm (CASUAL)	
	Pseudonym	Profession	*n* accent-1 long vowels	*n* accent-1 short vowels	*n* accent-2 long vowels	*n* accent-2 short vowels	*n* unstressed long vowels	*n* unstressed short vowels	*n* phrase-final vowels	*n* /r/ in coda position	mean speech rate in milliseconds	mean lexical frequency	age	racialized binary	Social class (PC1)	SEI	monthly income (1000 SEK)	formal education	parental SEI	parental formal education	*Mosaic* taste	*n* vocalic intervals	mean nPVIV (CASUAL)
'invandrare' social class working → upper	Reman	social worker	55	37	12	29	159	254	26	29	150	7 641	24	invandrare	0	24	25	33	0	1	lowbrow	340	55.8
	Malik	cook at cafeteria, street food	77	55	18	34	117	253	48	7	147	7 019	28	invandrare	1	24	25	33	0	1	lowbrow	333	48.0
	Max	youth worker	149	86	44	71	318	565	57	62	134	7 004	28	invandrare	4	25	15	50	0	17	lowbrow	742	48.8
	Antonio	caretaker, ex drug dealer	166	96	73	124	307	659	63	65	160	5 927	24	invandrare	8	11	25	17	0	58	lowbrow	849	45.6
	Hayder	caretaker, bouncer	85	59	38	77	224	403	36	24	160	6 243	25	invandrare	19	23	25	33	0	58	lowbrow	539	54.8
	Dawit	group home assistant	195	82	73	93	361	761	78	56	141	5 583	26	invandrare	20	27	15	67	0	42	lowbrow	961	51.3
	Solomon	youth worker	150	120	60	112	370	820	61	74	137	7 776	37	invandrare	20	25	35	67	0	33	lowbrow	993	59.0
	Murad	group manager, furniture retail	98	64	36	58	180	389	43	14	137	5 190	25	invandrare	22	44	25	33	0	50	lowbrow	504	45.2
	Abel	photographer, personal trainer	110	53	30	61	201	383	52	34	149	7 800	26	invandrare	25	51	25	33	0	33	lowbrow	500	54.6
	Mateo	waiter, bartender	93	63	33	48	278	432	60	43	145	7 502	36	invandrare	32	25	25	33	0	75	lowbrow	585	60.6
	Reza	waiter, napropathy student	120	55	46	39	230	457	39	28	154	5 634	25	invandrare	35	25	15	50	0	75	lowbrow	583	51.9
	Mezdar	youth worker (certified)	118	45	21	51	185	369	47	46	150	6 893	28	invandrare	38	39	45	50	0	33	highbrow	473	52.4
	Jocke	athletic coach, personal trainer	194	100	62	100	450	826	90	70	144	7 604	28	invandrare	42	51	15	67	0	58	lowbrow	1 065	57.0
	Sohrab	web architect	149	67	37	61	298	548	67	48	151	6 529	33	invandrare	49	75	45	83	0	42	lowbrow	714	63.5
	Shorty	social worker	89	82	22	45	195	325	62	30	137	8 545	26	invandrare	51	53	35	67	0	58	lowbrow	455	52.1
	Kevin	organisational consultant	137	67	40	64	316	597	54	42	140	7 532	39	invandrare	64	70	35	83	0	42	highbrow	761	58.1
	Tarik	field assistant; athletic coach	108	89	37	56	200	454	51	39	151	5 168	31	invandrare	73	51	85	33	0	67	highbrow	573	49.2
	Majeed	doctor	133	85	32	55	278	544	59	36	134	7 952	30	invandrare	76	89	45	83	0	75	lowbrow	693	65.4
	Parviz	business owner, physician	116	40	33	54	253	349	45	34	148	8 989	24	invandrare	91	79	25	83	0	83	highbrow	507	57.5
'svensk' social class working → upper	Per	demolition operative	66	28	27	36	205	363	43	27	153	9 555	29	svensk	15	15	25	17	0	33	lowbrow	453	78.3
	Rasmus	chef at major hotel	74	48	39	51	213	361	57	25	141	7 702	34	svensk	24	50	35	17	0	25	lowbrow	478	72.3
	Johnny	waiter, bartender	115	88	51	58	251	443	63	47	166	6 541	30	svensk	27	25	25	33	0	25	highbrow	591	64.6
	Sven	team leader, security firm	50	38	29	36	135	324	29	38	151	8 703	24	svensk	35	51	25	33	0	58	lowbrow	379	61.5
	Paul	self-employed, renovations	91	67	32	46	240	414	62	54	156	7 528	39	svensk	38	38	25	33	0	25	highbrow	543	74.3
	Jesper	waiter	61	49	35	64	225	427	45	38	132	7 672	23	svensk	39	25	5	50	0	50	highbrow	541	64.8
	Richie	pest exterminator	112	77	50	44	284	486	37	57	149	9 353	30	svensk	53	56	15	50	0	50	highbrow	637	60.4
	Thomas	middle office associate, hedgefund	173	68	35	65	376	648	54	62	140	8 233	32	svensk	61	57	45	33	0	50	highbrow	822	65.6
	Nils	supermarket clerk	100	62	23	51	208	456	44	34	152	8 233	39	svensk	64	29	25	33	0	92	highbrow	565	57.2
	August	section manager, IT	66	58	29	33	251	352	30	39	149	8 025	42	svensk	74	73	65	67	0	42	highbrow	482	67.0
	Jonte	middle school teacher, special ed	113	41	55	60	263	513	47	34	144	6 249	43	svensk	77	71	45	83	0	42	highbrow	640	71.2
	Martin	massage therapist, former teacher	101	37	38	31	287	422	41	23	148	8 346	39	svensk	82	67	25	83	0	75	highbrow	557	91.0
	Gunnar	high school teacher	118	51	30	43	235	390	52	38	144	7 395	37	svensk	82	82	45	83	0	58	highbrow	517	60.0
	Joseph	communications strategist/PR	84	47	24	37	148	296	38	23	178	7 675	33	svensk	84	75	45	33	0	92	highbrow	378	56.0
	Jan-Bertel	portfolio manager, hedgefund	126	65	39	44	292	433	69	28	132	6 478	31	svensk	87	72	95	83	0	42	highbrow	584	61.6
	Johan	chief marketing officer	38	17	29	34	82	173	20	22	162	9 098	39	svensk	89	72	35	67	0	83	highbrow	220	57.5
	Jan-Axel	trader, hedgefund	92	48	45	45	154	347	61	35	170	7 633	32	svensk	100	72	100	83	0	50	highbrow	443	51.7
		Means:	*109*	*62*	*38*	*56*	*244*	*451*	*51*	*39*	*148*	*7 415*										*21 000*	57.8
		Percentages:	19%	11%	6%	10%	42%	77%	9%	7%													

Table 6.10: Overview of dataset for internal and social predictors. Internal predictors: total counts for accent 1 and 2 long and short vowels, unstressed long and short vowels, vowels in phrase-final position, rhotics in coda position, mean speech rate in milliseconds, mean lexical frequency. Social predictors: age, racialization, Social class (PC1), SEI, monthly income (1000 SEK), education, parental SEI, parental education, and *Mosaic* market-segmentation taste. The final two columns are the *n* vocalic intervals counted for nPVIV in CASUAL speech and the mean nPVIV in CASUAL speech. Speakers are grouped by the racialized *invandrare*/*svensk* binary and ranked according to Social class (PC1).

	PC1	PC2	PC3
Standard deviation	1.71	1.11	0.89
Proportion of Variance	0.49	0.20	0.13
Cumulative Proportion	0.49	0.69	0.82

	PC1	PC2	PC3
sei	−0.51	−0.22	0.24
income	−0.36	−0.43	−0.45
education	−0.39	−0.39	0.60
parental_education	−0.35	0.61	0.27
parental_sei	−0.42	0.50	−0.16
mosaic_taste	−0.41	−0.02	−0.53

Table 6.11: Principle Components Analysis (PCA) of the six social-class predictors: first three components (82% of total variance).

	Model 1 *Internal only*		*Best* ↓ Model 2 *Social class (PC1)* *age + race : class*		Model 2′ *Social class (PC1)* *age : race + class*		Model 2″ *Social class (PC1)* *age : class + race*	
Response variable	nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)	
(Intercept)	7.8 (4.7)$^{\circ}$		2.3 (9.5)		1.0 (12.8)		−20.4 (17.8)	
Internal predictors		vif		vif		vif		vif
accent_1_long·*yes*	23.6 (1.0)***	1.4	23.7 (1.0)***	1.4	23.7 (1.0)***	1.4	23.7 (1.0)***	1.4
accent_1_short·*yes*	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2
accent_2_long·*yes*	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2
accent_2_short·*yes*	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2
unstressed_long·*yes*	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8
unstressed_short·*yes*	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7
coda_r·*yes*	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0
phrase_final·*yes*	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0
log(lexical_freq)	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1
log(speech_rate)	6.0 (0.8)***	1.1	5.9 (0.8)***	1.1	5.9 (0.8)***	1.1	5.9 (0.8)***	1.1
Social predictors				vif		vif		vif
age	-		0.7 (0.3)**	1.5	0.5 (0.4)	2.4	1.2 (0.6)*	6.5
racialization·*invandrare*	-		−23.0 (5.3)***	4.8	−18.4 (16.9)	38.7	−9.9 (3.2)**	1.4
pc1_social_class	-		−0.2 (0.1)**	3.1	−0.0 (0.1)	1.4	0.3 (0.3)	41.7
racialization·*invandrare*:pc1_social_class	-		0.3 (0.1)**	4.1	-		-	
age:racialization·*invandrare*	-		-		0.3 (0.5)	34.2	-	
age:pc1_social_class	-		-		-		−0.0 (0.0)	61.8
aic	218 859.3		218 844.6		218 849.3		218 856.2	
bic	218 970.6		218 987.7		218 992.4		218 999.3	
Log Likelihood	−109 415.6		−109 404.3		−109 406.6		−109 410.1	
Num. obs.	21 000		21 000		21 000		21 000	
Num. groups: vowel	3 099		3 099		3 099		3 099	
Num. groups: speaker	36		36		36		36	
Var: vowel (Intercept)	70.1		70.0		70.1		70.0	
Var: speaker (Intercept)	96.3		47.8		62.5		60.4	
Var: Residual	1 906.1		1 906.1		1 906.1		1 906.1	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Table 6.12: Three mixed-effects regression models that test the interaction of age, race, and social class in different combinations. They are positioned next to the internal model for comparison. Model 2 is determined to have the best fit. Models 2′ and 2″ have inferior fits.

more fundamental set of independent variables exists, perhaps fewer in number than the x's, which determine the value the x's will take. [These] have been called mental factors in recent psychological literature. (Hotelling 1933:1)

By the same logic, one can think of SEI, income, education, parental SEI, parental education, and Mosaic taste as tangible, measurable expressions of more latent and universal factors. One such factor is likely social class.

Using *prcomp* in *R*, I conducted a Principle Components Analysis (PCA) for the six social-class predictors *SEI*, *income*, *education*, *parental SEI*, *parental education*, and *Mosaic taste*. The pattern matrix was rotated orthogonally. The analysis rendered six principle components, the first three of which are shown in Table 6.11. They constitute 82% of the total variance. Principle component 1 (PC1) accounts for 49% of the total variance, and all of the six predictors contribute substantially with the absolute value of factor coefficients ranging between 0.35 and 0.51. I interpret PC1 to be an operationalization of current social class for a number reasons. First, it is the strongest bind for six artifacts that are qualitatively motivated in my literature review on social class. Second, the factor coefficients weigh all in the same direction, implying that an increase in each will contribute to the principle component in the same direction. Third, PC2 only accounts for 20% of the total variance and has factor coefficients SEI, income, and education that go in the opposite direction of parental education and parental SEI[10]. Logically, we know that not to be true for social class.

I extracted the principle-components data frame for PC1 and converted it to a 100-point scale. The converted PC1 scores are provided in Table 6.10 (p. 151) under the heading SOCIAL CLASS (PC1).

Since the research question concerns the speech of young racialized working-class speakers, the following three combinations are possible in a statistical model[11]:

$$\begin{aligned} \rightarrow\ & age + race : class \\ & age : race + class \\ & age : class + race \end{aligned}$$

The general rule of thumb is ten participants per parameter. According to Heo and Leon (2010), a fourfold increase is needed to in order to run three-way interactions ($n = 80$). I have therefore attempted to limit single-style models to two-way interactions (unless the analysis is exploratory as in Chapter 9, p. 269).

The class predictor was PC1_SOCIAL_CLASS. I ran mixed-effects linear regression models in *lme4* in *R* with the following three calls:

[10] PC2 can perhaps be interpreted as an operationalization of the portion of one's class background that is *not* represented in one's current social class. In this sense, it could also be seen as an operationalization of *social mobility*. While interesting for perhaps other questions, PC2 is not used in this analysis. PC3 has an Eigenvalue < 1 ($stdev^2 = 0.89^2 = 0.79$) and is therefore discarded.

[11] The symbol : is the annotation in *R* for a statistical interaction.

1. nPVIV ~ accent_1_long + accent_1_short + accent_2_long + accent_2_short + unstressed_long + unstressed_short + coda_r + phrase_final + log(lexical_freq) + log(speech_rate) + age + racialization*pci_social_class + (1|speaker) + (1|vowel)

2. nPVIV ~ accent_1_long + accent_1_short + accent_2_long + accent_2_short + unstressed_long + unstressed_short + coda_r + phrase_final + log(lexical_freq) + log(speech_rate) + age * racialization + pci_social_class + (1|speaker) + (1|vowel)

3. nPVIV ~ accent_1_long + accent_1_short + accent_2_long + accent_2_short + unstressed_long + unstressed_short + coda_r + phrase_final + log(lexical_freq) + log(speech_rate) + age * pci_social_class + racialization + (1|speaker) + (1|vowel)

Variance inflation factors (VIF) were calculated separately using the *car* package. The three models are shown in Table 6.12 alongside a reproduction of the internal-only model from Table 6.6 (p. 142).

Model 2 is marginally better than the other two social models because of its lower AIC and BIC. Model 2 also has the overall lowest variance inflation factors (VIF) of the three social models. In fact, the VIF values for the interactions in Models 2′ and 2″ are unacceptably high. All of the social models (Models 2, 2′, 2″) show a mild improvement in AIC over the internal-only model (Model 1). The BIC, however, is worse because of the higher parameter penalty. An increase to 14 predictors in the social model from the 10 predictors in the internal-only model renders an AIC penalty of 8 ($2 \cdot 4 = 8$) and a BIC penalty of 40 ($ln(23\,000) \cdot 4 = 10 \cdot 4 = 40$). Regardless of penalty preference, the log likelihoods of all the social models show improvement over the internal-only model.

The preferred model 2 shows strong and significant effects for age, race, social class, and the interaction of *invandrare* identity and social class.

Age In Model 2, the coefficient for age is significant and at 0.7, the amount by which the model predicts nPVIV to increase (become *less staccato*) for every year older the speaker is. In other words, it predicts that a 10-year increase in age will increase one's nPVIV by 7 (*less staccato* by 7). The right-side panel in Figure 6.5 visualizes the data and shows that age has a stable and mild effect for both *invandrare* and *svensk* speakers with *invandrare* speakers in the lead.

Racialization In Model 2, the coefficient for racialization is significant and at −23.0, the amount by which the model predicts nPVIV to decrease (become *more staccato*) if the speaker identifies as *invandrare* (although this main effect must always be interpreted alongside class because an interaction effect with class is simultaneously present). Both the left-side and right-side panels in Figure 6.5 demonstrates that *invandrare* speakers have overall lower nPVIV than *svensk* speakers, except for the upper-class speakers.

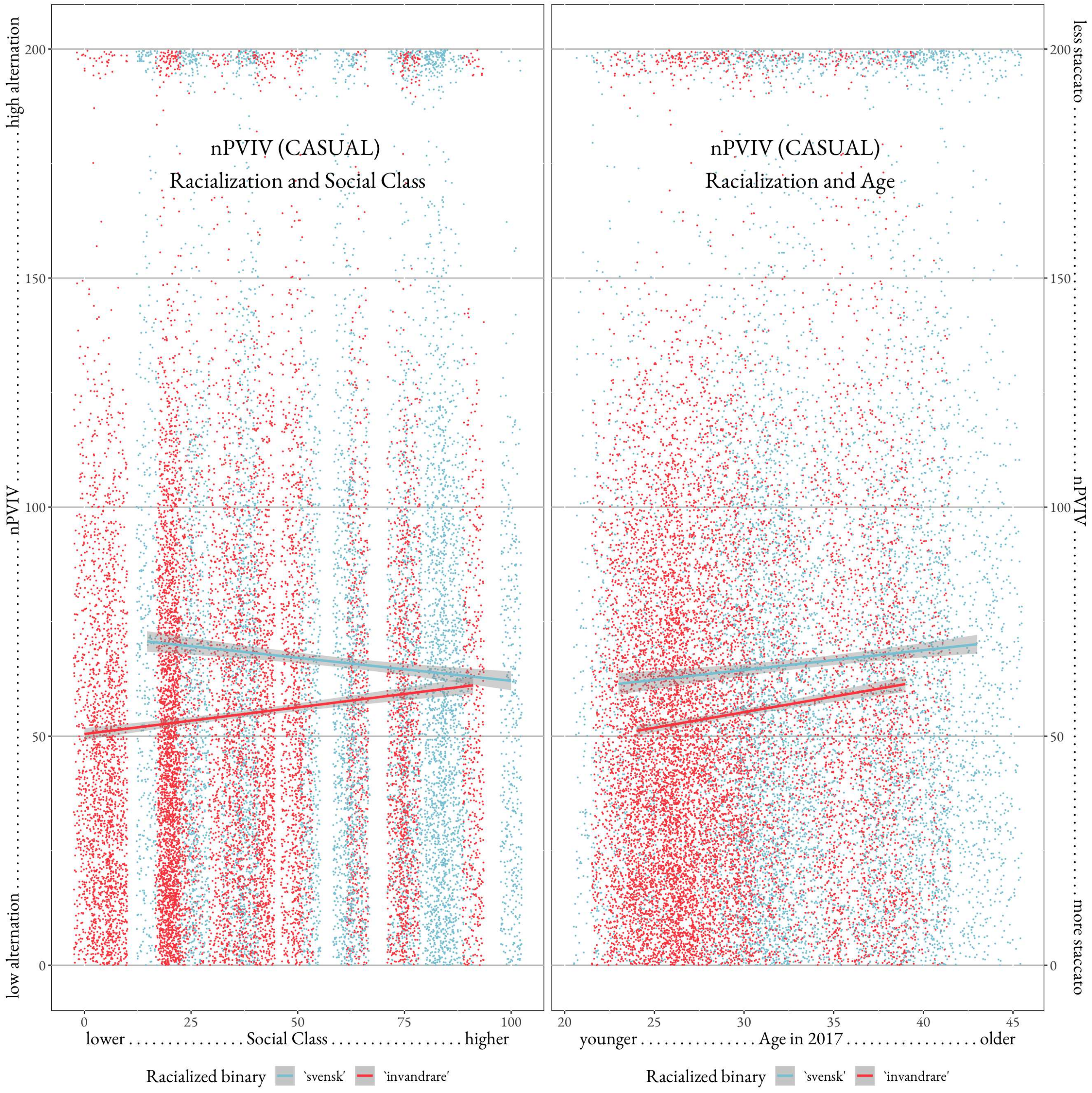


Figure 6.5: Scatterplots of RACIALIZATION and SOCIAL CLASS against nPVIV (left) and racialization and AGE against nPVIV (right). SOCIAL CLASS is defined as the first principle component (PC1) of the Principle Components Analysis of the six social factors *SEI*, *income*, *education*, *parental SEI*, *parental education* and *Mosaic taste*. Least-squares regression lines are plotted with confidence-interval halos.

Social class In Model 2, the coefficient for social class is significant and at −0.2, the (absolute) amount by which the model predicts nPVIV to decrease (become *more staccato*) for every point increase in social class if the speaker identifies as *svensk*. This is reversed if a speaker identifies as *invandrare*. The coefficient for the interaction [*invandrare* identity : social class] is significant and at 0.03, the amount by which the model predicts nPVIV to increase (become *less staccato*) for every point increase in social class if the speaker identifies as *invandrare*. The left-side panel in Figure 6.5 visualizes the data by plotting social class and racialization against nPVIV. It shows that as social class increases, the difference in rhythm between *svensk* and *invandrare* speakers decreases.

Below is a hypothetical case study of four speakers from each pole of the racial and class cline. They are aged 30 and are each producing a long accent-1 vowel next to a short unstressed vowel at an average rate of speech (150 ms syllables) in an average-frequency word (7000):

Control calculation	→ 2.3 + 23.7 + 15.5	−(0.3 · *ln*(7000))	+(5.9 · *ln*(150))	+(0.7 · 30)
	→ 2.3 + 23.7 + 15.5	−2.7	+30.0	+21
	→ 89.8			

lower-class *invandrare* (*social class* = 20)	→ 89.8 − 23.0 − (0.2 · 20) + (0.3 · 20)	= 68.8	→ *most 'staccato'*
lower-class *svensk* (*social class* = 20)	→ 89.8 − (0.2 · 20)	= 85.8	→ *least 'staccato'*
higher-class *invandrare* (*social class* = 80)	→ 89.8 − 23.0 − (0.2 · 80) + (0.3 · 80)	= 74.8	→ *inbetween*
higher-class *svensk* (*social class* = 80)	→ 89.8 − (0.2 · 80)	= 73.8	→ *inbetween*

According to Model 2, the difference in nPVIV lies between *invandrare* and *svensk* on the lower side of the class cline. Lower-class *invandrare* are predicted to have the most staccato rhythm (e.g., 68.8). An unexpected finding is that lower-class *svensk* speakers are predicted to have the least staccato rhythm (e.g., 85.8), and higher-class speakers are predicted to have an intermediate rhythm. The *invandrare/svensk* difference in rhythm is predicted to disappear within the higher classes (e.g., 73.8 and 74.8).

Evaluating each class component in isolation

I am not of the opinion that a single composite measure should be the only predictor analyzed. A more robust approach is one that examines all social predictors in isolation to see whether this general trend replicates itself. Separate mixed-effects linear regression models were run according to the schema of Model 2 (Figure 6.12, *age+race* : *class*) for each of the following social class parameters: SEI, INCOME, EDUCATION, PARENTAL_SEI, PARENTAL_EDUCATION, and MOSAIC_TASTE. I used *lme4* in *R* with the following call:

> nPVIV ~ ACCENT_1_LONG + ACCENT_1_SHORT + ACCENT_2_LONG + ACCENT_2_SHORT + UNSTRESSED_LONG + UNSTRESSED_SHORT + CODA_R + PHRASE_FINAL + LOG(LEXICAL_FREQ) + LOG(SPEECH_RATE) + AGE + RACIALIZATION* [*individual class parameter*] + (1|SPEAKER) + (1|VOWEL).

Variance inflation factors (VIF) were calculated separately using the *car* package.

	Model 1 *Internal only*		Model 2 *Social class (PC1)*		Model 3 *SEI*		Model 4 *Income*		Model 5 *Education*		Model 6 *Parental SEI*		Model 7 *Parental education*		Model 8 *Mosaic taste*	
Response variable	nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)	
Internal predictors		VIF		VIF		VIF		VIF		VIF		VIF		VIF		VIF
(Intercept)	7.8 (4.7)$^{\circ}$		2.3 (9.5)		3.5 (9.9)		0.5 (10.1)		1.9 (10.2)		19.3 (15.2)		3.6 (10.5)		−0.3 (10.1)	
accent_1_long·*yes*	23.6 (1.0)***	1.4	23.7 (1.0)***	1.4	23.7 (1.0)***	1.5	23.7 (1.0)***	1.4	23.7 (1.0)***	1.5	23.6 (1.0)***	1.4	23.6 (1.0)***	1.4	23.7 (1.0)***	1.4
accent_1_short·*yes*	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2
accent_2_long·*yes*	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2
accent_2_short·*yes*	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2	10.8 (1.2)***	1.2	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2
unstressed_long·*yes*	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8
unstressed_short·*yes*	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7
coda_r·*yes*	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0
phrase_final·*yes*	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0	7.0 (1.1)***	1.0	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0
log(lexical_freq)	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1
log(speech_rate)	6.0 (0.8)***	1.1	5.9 (0.8)***	1.1	5.9 (0.8)***	1.1	5.9 (0.8)***	1.1	5.9 (0.8)***	1.1	5.9 (0.8)***	1.1	6.0 (0.8)***	1.1	5.9 (0.8)***	1.1
Social predictors				VIF		VIF		VIF		VIF		VIF		VIF		VIF
age	-		0.7 (0.3)**	1.5	0.6 (0.3)*	1.5	0.6 (0.3)*	1.4	0.5 (0.3)$^{\circ}$	1.5	0.6 (0.3)*	1.3	0.6 (0.3)*	1.3	0.7 (0.3)*	1.4
racialization·*invandrare*	-		−23.0 (5.3)***	4.8	−24.1 (6.4)***	6.6	−14.3 (5.4)**	4.5	−20.2 (6.7)**	6.6	−38.4 (14.0)**	29.9	−20.0 (6.5)**	6.3	−16.1 (4.9)**	3.6
pc1_social_class	-		−0.2 (0.1)**	3.1	-		-		-		-		-		-	
racialization·*invandrare*:pc1_social_class	-		0.3 (0.1)**	4.1	-		-		-		-		-		-	
sei	-		-		−0.2 (0.1)*	2.8	-		-		-		-		-	
racialization·*invandrare*:sei	-		-		0.3 (0.1)**	6.6	-		-		-		-		-	
income	-		-		-		−0.2 (0.1)*	1.6	-		-		-		-	
racialization·*invandrare*:income	-		-		-		0.2 (0.1)	4.0	-		-		-		-	
education	-		-		-		-		−0.1 (0.1)	2.1	-		-		-	
racialization·*invandrare*:education	-		-		-		-		0.2 (0.1)$^{\circ}$	7.3	-		-		-	
parental_sei	-		-		-		-		-		−0.4 (0.2)*	10.8	-		-	
racialization·*invandrare*:parental_sei	-		-		-		-		-		0.5 (0.2)*	23.1	-		-	
parental_education	-		-		-		-		-		-		−0.2 (0.1)$^{\circ}$	2.4	-	
racialization·*invandrare*:parental_education	-		-		-		-		-		-		0.2 (0.1)$^{\circ}$	7.1	-	
mosaic_taste·*highbrow*	-		-		-		-		-		-		-		−10.6 (5.2)*	4.1
racialization·*invandrare*:mosaic_taste·*highbrow*	-		-		-		-		-		-		-		9.0 (6.7)	2.6
aic	218 859.3		218 844.6		218 846.1		218 847.6		218 849.4		218 846.9		218 849.2		218 832.9	
bic	218 970.6		218 987.7		218 989.3		218 990.7		218 992.5		218 990.1		218 992.4		218 976.0	
Log Likelihood	−109 415.6		−109 404.3		−109 405.1		−109 405.8		−109 406.7		−109 405.5		−109 406.6		−109 398.4	
Num. obs.	21 000		21 000		21 000		21 000		21 000		21 000		21 000		21 000	
Num. groups: vowel	3 099		3 099		3 099		3 099		3 099		3 099		3 099		3 099	
Num. groups: speaker	36		36		36		36		36		36		36		36	
Var: vowel (Intercept)	70.1		70.0		70.1		70.1		70.1		70.0		70.1		70.0	
Var: speaker (Intercept)	96.3		47.8		51.5		54.4		57.0		55.3		56.8		56.4	
Var: Residual	1 906.1		1 906.1		1 906.1		1 906.1		1 906.1		1 906.1		1 906.1		1 906.1	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Table 6.13: Eight separate mixed-effects linear regression models. All models contain the same internal predictors. Model 1 contains only internal predictors. Models 2–8 contain the social predictors *age* and the *racialized binary* along with different operationalizations of social class: Model 2 - Principle component 1 of the six class measurements, Model 3 - *SEI*, Model 4 - *Income*, Model 5 - *Education*, Model 6 - *Parental SEI*, Model 7 - *Parental education*, Model 8 - *Mosaic taste*. For categorical predictors, the reference category is in italics. Coefficients are indicated in the center column, standard errors in the parentheses, and variance inflation factors (VIF) to the right.

Table 6.13 contains the six new models next to the internal-only Model 1 and the social-class (PC1) Model 2. The regression outputs show that all models maintain similar coefficients and errors for the internal predictors. All models the show same directionality for age and racialization, albeit with varying coefficient strengths. For each class predictor, all models show the same directionality, albeit with varying intensities and significance levels. Lower class predicts a higher nPVIV (*less staccato*) for *svensk* speakers and lower nPVIV (*more staccato*) for *invandrare* speakers; higher class attenuates these effects. I will explain each predictor effect in more detail below.

Age In models 2 through 8, the coefficient for age is significant and ranges between 0.5 and 0.7, the amount by which the model predicts nPVIV to increase (become *less staccato*) for every year older the speaker is. In other words, if speaker A is 10 years older than speaker B, and all other factors are constant, A's nPVIV is expected to be between 5 and 7 points higher than B's. The right-side panel in Figure 6.5 visualizes the data and shows that age is a stable and mild effect for both *invandrare* and *svensk* speakers with *invandrare* speakers in the lead.

Racialization In models 2 through 8, the coefficient for racialization fluctuates between -14.3 and -38.4, the amount by which the model predicts nPVIV to decrease (become *more staccato*) if the speaker identifies as *invandrare*. Both the left-side and right-side panels in Figure 6.5 demonstrates that *invandrare* speakers have overall lower nPVIV than *svensk* speakers, except for the speakers from the higher classes. This latter interaction is discussed in the below sections.

Occupational status (SEI) Model 3 mirrors Model 2 almost exactly. This is unsurprising because SEI weighs the strongest for Principle component 1 (See Table 6.12). The model predicts that each one-point increase in SEI (100-point scale) will reduce nPVIV by -0.2 (becoming *more staccato*). However, if a speaker identifies as *invandrare*, the reverse is predicted; every point increase in SEI (100-point scale) is predicted to increase nPVIV by 0.3 (becoming *less staccato*). To illustrate what the model predicts on a more general level, the following is a case study for four hypothetical speakers, aged 30, from two poles of the racialization and class spectrum. Let us say they are each producing a long accent-1 vowel next to a short unstressed vowel at an average rate of speech (150 ms syllables) in an average-frequency word (7000):

low-SEI *invandrare* (SEI= 20)	$\rightarrow 88 - 24.1 - (0.2 \cdot 20) + (0.3 \cdot 20)$	$= 65.9$	$\rightarrow$ *most 'staccato'*
low-SEI *svensk* (SEI= 20)	$\rightarrow 88 - (0.2 \cdot 20)$	$= 84.0$	$\rightarrow$ *least 'staccato'*
high-SEI *invandrare* (SEI= 80)	$\rightarrow 88 - 24.1 - (0.2 \cdot 80) + (0.3 \cdot 80)$	$= 71.9$	$\rightarrow$ *inbetween*
high-SEI *svensk* (SEI= 80)	$\rightarrow 88 - (0.2 \cdot 80)$	$= 72.0$	$\rightarrow$ *inbetween*

The stratification predicted here for SEI resembles that of *Social class (PC1)*. The *invandrare* working class is predicted to have the lowest nPVIV (65.9, *most staccato*), and the *svensk* working class – the highest (84.0, *least staccato*). The elites are predicted to have an intermediate-level nPVIV (71.9–72.0).

Income Model 4 shows similar trends to Models 2 and 3, but the interaction with racialization is non-significant.

Formal education Model 5 shows similar trends to Models 2, 3, and 4; yet, education has no statistical significance. This could be due to the fact that some high-status participants do not have more than a Bachelor's degree and that some low-status speakers do have Bachelor's degrees. The index also does not capture the difference in prestige between attending an established research university like Uppsala or Lund versus a newer teaching university like Södertörn or Malmö Högskola. The direction of the coefficients for education, however, resemble Model 2, 3, and 4.

Parental occupational status Model 6 shows similar trends to Models 2, 3, 4, and 5 with a significant effect for both parental occupational status and its interaction with racialization. Its coefficients are the strongest among all of the (100-point) models, indicating that it has the greatest predictive power on nPVIV[12]. The trend resembles Models 2 and 3: higher parental occupational status would predict opposite effects for *svensk* and *invandrare* speakers. The model predicts that each one-point increase in one's parent's SEI (100-point scale) will reduce nPVIV by -0.4 (becoming *more staccato*). However, if a speaker identifies as *invandrare*, the reverse is predicted; every point increase in one's parent's SEI (100-point scale) is predicted to increase nPVIV by 0.5 (becoming *less staccato*). To tangibly illustrate what the model predicts, the following is another case study for four hypothetical speakers, aged 30, from two poles of the racialization and class spectrum. Let us say they are each producing a long accent-1 vowel next to a short unstressed vowel at an average rate of speech (150 ms syllables) in an average-frequency word (7000):

Invandrare with low-SEI parent (SEI= 20)	$\rightarrow 109.1 - 38.4 - (0.4 \cdot 20) + (0.5 \cdot 20)$	$= 52.7$	$\rightarrow$ *most 'staccato'*
Svensk with low-SEI parent (SEI= 20)	$\rightarrow 109.1 - (0.4 \cdot 20)$	$= 101.1$	$\rightarrow$ *least 'staccato'*
Invandrare with high-SEI parent (SEI= 80)	$\rightarrow 109.1 - 38.4 - (0.4 \cdot 80) + (0.5 \cdot 80)$	$= 78.7$	$\rightarrow$ *inbetween*
Svensk with high-SEI parent (SEI= 80)	$\rightarrow 109.1 - (0.4 \cdot 80)$	$= 77.1$	$\rightarrow$ *inbetween*

The case study shows that parental SEI predicts a larger divide in nPVIV between the two working classes than SEI or social class PC1. One way to interpret this is that class background plays a stronger role in the vernacular speech of adults than would current social class or occupational status. This sample of speakers has a significant amount of social mobility as far as occupational status is concerned; the correlation coefficient between SEI and Parental SEI is only 0.44.

Parental formal education Model 7 shows similar trends to Models 2, 3, 4, 5, and 6, yet parental formal education has no statistical significance. This may be because there are a number of *invandrare* speakers whose parents have low-status jobs but advanced degrees from their home countries. Reza's father has a dentistry degree from his country of origin

[12] NB, however, that the standard error doubles for PARENTAL_SEI and RACIALIZATION·*invandrare*:PARENTAL_SEI.

and now drives a taxi. Antonio's father has an advanced engineering degree from his country of origin and now works as an auto mechanic. Sohrab's mother has a Master's degree from an elite university in her country of origin and now works as a home-care assistant. The coefficients, however, for parental education resemble those of Model 2 and 3: higher parental formal education would predict opposite effects for *svensk* and *invandrare* speakers.

Taste status   Model 8 shows similar trends to Models 2, 3, 4, 5, 6, and 7, with a significant effect for *Mosaic* taste. However, the interaction with racialization is non-significant. Nonetheless, the coefficients show the same trend of higher status predicting lower nPVIV for *svensk* speakers and higher nPVIV for *invandrare* speakers.

3.4 Section summary

The results in the preceding analysis can be summarized in the form of six main points.

1. Age has a mild predictive effect with consistently strong significance in every model run. Younger speakers are predicted to have lower nPVIV values (*more staccato*) than older speakers.

2. Racialization has a strong predictive effect with consistently strong significance in every model run. *Invandrare* speakers correspond with lower nPVIV values (*more staccato*), and *svensk* speakers with higher nPVIV values (*less staccato*). But the racialization effect is conditioned by social class, so it cannot be discussed in isolation. Rather, the higher the social class, the less racialization has an effect; see 5.

3. Based on some of the models, lower-class *svensk* speakers are predicted to have the highest nPVIV values (*less staccato*) in their vernacular speech.

4. Based on most of the models, lower-class *invandrare* speakers are predicted to have the lowest nPVIV values (*most staccato*) in their vernacular speech.

5. Based on some of the models, higher-class speakers, regardless of racialization, are predicted to have intermediate-level nPVIV values in their vernacular speech.

6. The class-related predictor with the strongest stratifying effect on nPVIV is parental SEI, which implies the primacy of embodied class on language practice versus measurements of less embodied class (i.e., SEI, income, formal education).

Returning to the original research question, *Does the speech of Stockholm's racialized working class have lower rhythmic alternation than other varieties in the city; i.e., is it 'staccato'?*, the evidence points to the affirmative. Further, some unintended findings emerged from this analysis. Lower-class *svensk* speakers seem to have a high-alternation speech rhythm that differs significantly from the rhythm of both lower-class *invandrare* and all higher-class speakers. Speech rhythm seems also to be changing in apparent time with lower alternation among younger speakers and higher alternation among older speakers.

4 Introducing network qualities to the model

The purpose of this analysis is to address the research question with an entirely different set of predictors – social network data – in order to see whether the same patterns emerge as did in the previous analysis.

If 'staccato' rhythm is a feature of multiethnolect, one would expect that speakers with lower nPVIV would have, let us say, a higher number of close contacts who are both working-class and *invandrare* than speakers with higher nPVIV. One might also expect that these close contacts speak in a manner that the speakers metapragmatically describe as 'ghetto' or 'hood' (Sw.: 'orten'). In this section, social network data will be examined in a similar analytic framework to the preceding section. This chapter will home in on the *qualities* of the participant networks and how well they predict the stratification of nPVIV.

4.1 Background: Network analysis in linguistics

Network analyses in linguistics typically model the qualities and frequency of a speaker's encounter with interlocutors in order to test questions about language variation and change. In such analyses, an actual list of personal contacts is harvested or, alternatively, proxies are made from institutional membership (e.g., schools in Dodsworth and Benton 2017). Network qualities are then operationalized by assigning values to the relationship the speaker has to the interlocutor, numerically representing some property (such as social class) of the interlocutor (or institution), and/or numerically representing the speaker's position within the field of interlocutors (i.e., central or peripheral). From these data, a sociogram can be built that models a limited number of structural and qualitative properties for a real-life aggregation of social interactions. Bloomfield (1933) describes in detail an ideal version of such a model.

> Every speaker's language, except for personal factors which we must here ignore, is a composite result of what he has heard other people say. Imagine a huge chart with a dot for every speaker in the community, and imagine that every time any speaker uttered a sentence, an arrow were drawn into the chart pointing from his dot to the dot representing each one of his hearers. At the end of a given period of time, say seventy years, this chart would show us the density of communication within the community. [...] If we wanted to explain the likeness or and unlikeness between various speakers in the community, or, what comes to the same thing, to predict the degree of likeness and unlikeness between various speakers, our first step would be to count and evaluate the arrows and series of arrows connecting their dots. (Bloomfield 1933:46–47)

There are various approaches to conducting the above analysis, but they all are based on the following assumption. When individuals speak to someone they are close to, like a sibling or a childhood friend, they engage in a form of linguistic accommodation whereby the phonetic and syntactic forms from both speakers align. The high repetition of this behavior – day in and day out – with multiple close contacts (who, themselves, also engage in the same behavior), produces an environment conducive to dialect *focusing* (the establishment of stable norms for variants), changes from below, and consistent evaluations of 'external' innovations (often negative). When individuals speak to someone they are not close to, like a teacher or a colleague, they engage in a form of linguistic accommodation whereby phonetic and syntactic forms align and dis-align, depending on the properties of that relationship such as power asymmetries and fondness or lack thereof. The varying repetition of this behavior – day in and day out – with different contacts with which one has little intimacy, produces an environment conducive to dialect diffusion and *leveling* (the reduction of numbers of variants), changes from above, and less consistent evaluations of innovations. In Dodsworth and Benton's (2017) words, "the convergence over time of dialects in contact is commonly believed to result from linguistic accommodation during repeated face-to-face contact (Auer & Hinskens 2005; Cindy, Ogay, & Giles 2005; Giles 1973; Giles & Phillip 1979; Trudgill 1986)" (Dodsworth & Benton 2017:373).

The use of sociograms, in essence, approximates the cognitive embodiment of cumulative speech-accommodation moments. It therefore requires one to make a number of crude assumptions, which include dramatically simplifying the complexities of real relationships. In fact, according to Bergs (2005:23–24), social networks are abstract metaphors that somehow represent social structures. In this sense, they are not so much empirically testable constructs as they are heuristic devices (2005:24). Eckert (2000) refers to them as representative "slices of social practice".

> Each sociogram represents a slice of social practice, which fits into the whole quite differently in different communities of practice. [...] Some people spend more time with their peers than others: some go out on weeknights with their friends, while others stay at home doing homework. Some friendship groups see each other only in school, while others spend every available moment together. And the nature of the friendship will determine the kinds of interactions people engage in: whether they share problems, get in trouble together, do homework ride motorcycles, hang out and talk, go places. It would be impossible to ask enough questions to control for this kind of difference in ties. Nonetheless, it is not unreasonable to assume that these kinds of differences can have an important influence on linguistic behavior. (Eckert 2000:175)

The use of networks in sociolinguistics has typically relied on *ego-star* sociograms; that is to say, miniature, isolated sociograms that model only the ties to one principle actor. According to Sharma (2017), ego-star analyses provide "some network information per individual while still permitting stratified sampling, as participants need not be linked in a single network" (2017:395). This dissertation makes use of stratified sampling and ego-star information. Similarly, the majority of sociolinguistic studies that have used networks have also made use of ego-star networks[13]. Unlike whole-network models, ego-star models are granular, which means their properties are easy to model with linear-regression analyses.

According to Sharma (2017), the sociolinguistic network analyses have typically either examined *structural* properties within each ego-star network or *qualities* of these networks.

> We can distinguish between *structural* and *content* (or *quality*) components at any level. Structural components represent the actual geometry or arrangement of ties, such as density. Content components involve the qualities present within these ties, for instance relating to interaction, such as multiplexity, reciprocity, and transactions (Bergs 2005; Milardo 1988), or to attributes, such as ethnicity of individuals. (Sharma 2017:396)

In her investigation of phonetic change across three cohorts of British Asians in London, Sharma (2017) found that network structure and quality each contributed to solving different scales of analysis. When it came to the question of which specific linguistic features were used by which groups (i.e., a micro scale), network quality offered the most explanatory power. For example, in the youngest cohort of speakers, use of postalveolar /t/ corresponded with having more Asians in one's social network. (Sharma 2017:412). When it came to generalizing about overall repertoires and language change over time (i.e., macro scale), network structure offered the most explanatory power. For example, her participants' accent range (an index based on the diversity of variants used for four variables across various speech domains) corresponded with having a network that was more fragmented in structure[14]. In line with Sharma's theoretical premise, the current chapter will focus predominantly on network qualities.

In Chapter 9 (p. 260) I will review the network-structural studies (Bergs 2005; Bortoni-Ricardo 1985; Cheshire 1982; Dodsworth & Benton 2017; Eckert 2000; Fagyal, Swarup, Escobar, Gasser, & Lakkaraju 2010; Labov 1972b, 1973; Milroy 1987) because its research question concerns rhythm in the context diffusion over time. The research question for this chapter, however, is relatively micro-level and concerns a specific feature (rhythm) and its placement within a specific community (young, racialized, working-class men). I will therefore only review the literature here that has concerned itself with network qualities.

One of the earliest studies on the effect of network quality on language was conducted by Blom and Gumperz (1972 [1986]) in the small arctic village of Hemnesberget in Norway. They found that residents with solely local ties did not style-shift to standard Bokmål Norwegian even when the topics were authoritative, political, or formal. On the other hand,

[13] The notable exceptions are Labov (1973), Bortoni-Ricardo (1985) and Eckert (2000), which I will review more in more detail in Chapter 9 (p. 260).

[14] Sharma (2017) refers to this as *network diversity*, which at first implies it might be a qualitative attribute. However, *diversity* refers to the degree of fragmentation within the network: "how diverse the distribution of ties in the network is across different groups—that is, how many different social worlds a person's network is composed of and how balanced they are in terms of relative size" 2017:401.

residents with external ties, such as university students, did style-shift according to topic. Among the village's elites, Bokmål was used extensively among close friends, and the village dialect was only used for jest. They concluded that the relationship between genre and shared prestige was not straightforward; individuals with hyperlocal networks kept to the vernacular regardless of genre (1972 [1986]:427–432).

Gal (1979) found that speakers selected Hungarian and German in adherence to a complex relationship to their peasant and wage-worker identities, respectively, as well as their ages (1979:159). The community-specific social class schema of 'peasantness' was operationalized by means of a seven-point index by which she measured each of the 32 participants (1979:136–140). She then developed a simplified index by which to operationalize the participants' weekly intelocutors in order to construct a class-based sociogram (1979:140–141). The network index had to be crude because participants possessed less information about their contacts than about themselves (and often were reluctant to disclose much about what they did know). Nonetheless, the social class of participant networks was a much stronger correlate of German-versus-Hungarian choice ($R = 0.78$) than the social class of the participants itself ($R = 0.67$) – despite the former being a more crude metric than the latter (1979:139, 141).

> Certainly people's habitual language choices are somehow related to their identities, to their position on the social dimension of peasant-to-Austrian urbanite. But their are several ways of analyzing this relationship. Each mode of analysis carries with it a conception of the relationship between linguistic and social facts. The most common approach is to choose an aspect of the speaker's social status such as age, class, or, in this case, degree of 'peasantness', and to devise a measure of it that can be correlated with linguistic patterns. [...] A second approach adds to the first by considering not the identity of the speaker, but the statuses of those people with whom the speaker most often interacts. This is a measure of social networks specifically tied to, rather than being independent of, the nature and frequency of the speaker's interaction with others. This view argues that speakers' linguistic behaviors are constrained and shaped by the sorts of social contacts they maintain [...]. (Gal 1979:131–132)

Ethnicity is one of the most-common qualities examined when investigating the distribution and/or stratification of a linguistic variable. Cheshire, Fox, Kerswill, and Torgersen (2008) investigated the correlation between use of 14 phonetic variables – many of them key features of London's multiethnolect – and the percentage of close friends of a different ethnicity among speakers. The analysis made a strong case for concluding that Anglo-speakers with large non-Anglo friendship groups were the "bridge for the transmission of minority ethnic features" into the general population (Cheshire et al. 2008:14). Li Wei (1994) developed a 'Chinese ethnic index' based on how many pre-migration Chinese ties the subject has in each of his three network types: *exchange*, *interactive* and *passive*. He found a strong correlation between Chinese use and this index. Matsumoto (2010) found that speaking Japanese within families on the island of Palau was highly correlated with having a high number of extra-familial Japanese ties. Meyerhoff and Schleef (2012) found that friendship network was the second-largest constraint on whether Polish teens produced the [in] variant of (ing) in Edinburgh. Teens with mostly Scottish friends were significantly less likely to produce the velar variant than those with mostly Polish friends. In their investigation of two British-Asian generations, Sharma and Sankaran (2011) found that the 'Asian-ness' of their speakers' networks only had a corresponding effect with Asian linguistic features for the older generation. For the younger generation, network had no significant effect, indicating that these features may have been repurposed to do other social work.

All of the above studies demonstrate that network qualities are robust tools for investigating and understanding the distribution and stratification of specific linguistic variables.

4.2 Operationalizing network qualities

For the present analysis on the effects of network qualities on nPVIV, two different datasets will be used. The first dataset is a series of proxy calculations about the 36 participants' networks based on interview questions about where they have lived and where they attended school. I estimated adolescent networks using school demographic data for the schools that the participants attended. Lifetime networks were calculated based on the demographic data of the neighborhoods the participants lived in throughout their lives. The second dataset comes from social-network interviews with 29 of the 36 participants about their current adult social networks (see p. 68). The qualities are as follows:

Network property (quality)	Actual or proxy network?	Current or childhood network?	Sample size	Description
Classmates - foreign background	proxy	childhood	$n = 36$	weighted average of the percentage of classmates with a 'foreign background' of every school the participant attended, grades 7–12
Classmates - working-class	proxy	childhood	$n = 36$	weighted average of the percentage of working-class classmates of every school the participant attended, grades 7–12
Lifetime neighborhood diversity	proxy	both	$n = 36$	weighted average of the percentage of residents with a 'foreign background' of every neighborhood the participant has lived in
Network - invandrare working-class	actual	adult	$n = 29$	percentage of close contacts, identified as *invandrare*, whose occupational SEI < 50
Network - svensk working-class	actual	adult	$n = 29$	percentage of close contacts, identified as *svensk*, whose occupational SEI < 50
Network - speaks 'hood'	actual	adult	$n = 29$	percentage of close contacts who the participant says speaks *hood* or *ghetto* or with a *suburban accent*

Proxy networks

This section introduces and describes the following predictors:

coding ID: CLASSMATES_FOREIGN_BACKGROUND
coding ID: CLASSMATES_WORKING_CLASS
coding ID: LIFETIME_NEIGHBORHOOD_DIVERSITY

The harvesting and modeling of a participant's actual childhood and lifetime networks is hardly feasible, but proxies can be built off of other data. In the main interview I asked the participant to list all the schools he had attended and for what grade (p. 66). I then catalogued a list of every school the participant attended by year along with a list of what secondary-school track he attended[15] by year. I then harvested demographic information on the schools from data provided by Skolverket (Eng. The Swedish National Agency for Education). Where available, I took demographic information specific to the year attended. Where data was missing, I used data available from the year that was closest to the missing year. I harvested two characteristics for each school: (1) percentage of students whose parents only had a high-school degree or less and (2) percentage of students of 'foreign background' as defined by Statistics Sweden (Statistics Sweden 2002; explained in more detail below.).

For *lågstadium* and *mellanstadium* (elementary school, grades 1-6), demographic data was not available for any year before 2009. However, for *högstadium* (middle school, grades 7–9), comprehensive data was typically available. I therefore decided to only use data from högstadium for the primary-school component of the index. For secondary schools, both foreign background and parental education data were available in detail for each specialty track for many years back. I calculated two indices from this: percentage of classmates with a working-class background in grades 7–12 (CLASSMATES_WORKING_CLASS) and percentage of classmates with a foreign background in grades 7–12 (CLASSMATES_INVANDRARE).

As an example, participant Malik attended högstadium (grades 7, 8, and 9) in Rinkebyskolan in years 2000, 2001, and 2002, respectively. He attended the construction-worker track at Stockholms Byggtekniska Gymnasium (formerly Kista Gymnasium) for secondary school in years 2003, 2004, and 2005. His school indices are calculated as follows:

	Rinkebyskolan			Stockholms Byggtekniska			
year	2000	2001	2002	2003	2004	2005	*mean*
Percent pupils of 'foreign background'	85.7	86.8	91.4	28.0	29.0	19.0	*56.7*
Percent pupils with parents with high-school education or less	90.9	90.9	83.6	73.0	74.0	75.0	*81.2*

The indices are a proxy for the demographics of his adolescent interlocutor pool: 56.7 percent of 'foreign background' and 81.2 percent working-class.

I also asked each participant to walk me through where he lived, at what age, and for how long since birth. I then collected data on the percentage of residents with a foreign background for each neighborhood for the most recent year that was available. This ranged between 2012 and 2018. *Foreign background* (*utländsk bakgrund*) is a prescribed statistical category in Sweden. Established by Statistics Sweden, it is defined as being either born abroad

[15] In Swedish secondary schools, one selects a specialty track such as *social science* or *carpentry* or *mathematics* and so on.

or having both parents born abroad. In other words, having a single parent born abroad does not count (Statistics Sweden 2002). I calculated a weighted average by multiplying the foreign-background statistics by the years lived in each neighborhood divided by the participant's age. For example, if a 26-year-old participant spent half his life in a neighborhood with 25% foreign-background residents and half his life in a neighborhood with 90% foreign-background residents, his 'lifetime neighborhood diversity' index would be 57.5:

$$\frac{(13 \cdot 25 + 13 \cdot 90)}{26} = 57.5$$

Figure 6.6 contains an annotated version of the same demographic map of greater Stockholm that was shown on page 28. The participants are annotated over the neighborhood where they currently live. Their lifetime neighborhood diversity score is provided adjacent to their pseudonyms. The map is color-coded according to the percentage of residents of 'foreign background' and shows striking segregation between the migrant-dense areas to the northwest and southwest on one hand and the more homogeneous north and east on the other hand. The participants' lifetime neighborhood diversity indices also reflect this with indices surpassing 80% in the core of the northwest and southwest suburbs.

Actual networks

This section introduces and describes the following predictors:

coding ID: NETWORK_INVANDRARE_WC
coding ID: NETWORK_SVENSK_WC
coding ID: NETWORK_SPEAKS_HOOD

Of the 36 participants in this study, 29 participated in an additional interview in which I investigated their current adult social networks (see p. 68). I asked the participants to think about whom they had spoken with over the last month, encouraging them to look through their phone and provide me with a comprehensive list. I entered the names into a spreadsheet and then returned to each name to ask about demographic data: *Was this contact 'very close', 'close', or 'not close'? What is his/her ethnicity? What does s/he do for a living?* For the latter question, I coded each occupation according to SEI (Ganzeboom & Treiman 2003). All the 'close' or 'very close' contacts who were identified as *invandrare* and had an SEI below 50 were tallied, and the percentage out of the total close contacts was calculated for the predictor 'Network - *invandrare* working-class' (NETWORK_INVANDRARE_WC). All the 'close' or 'very close' contacts who were identified as *svensk* and had an ISEI below 50 were tallied, and the percentage out of the total close contacts was calculated for the predictor 'Network - *svensk* working-class' (NETWORK_SVENSK_WC). I selected 50 because it constitutes the exact middle of the 100-point ISEI scale.

I also asked *What language do you use with him/her? How would you describe the speech of this contact?* For the latter question, answers like 'hood' ('orten'), 'suburban' ('förort'),

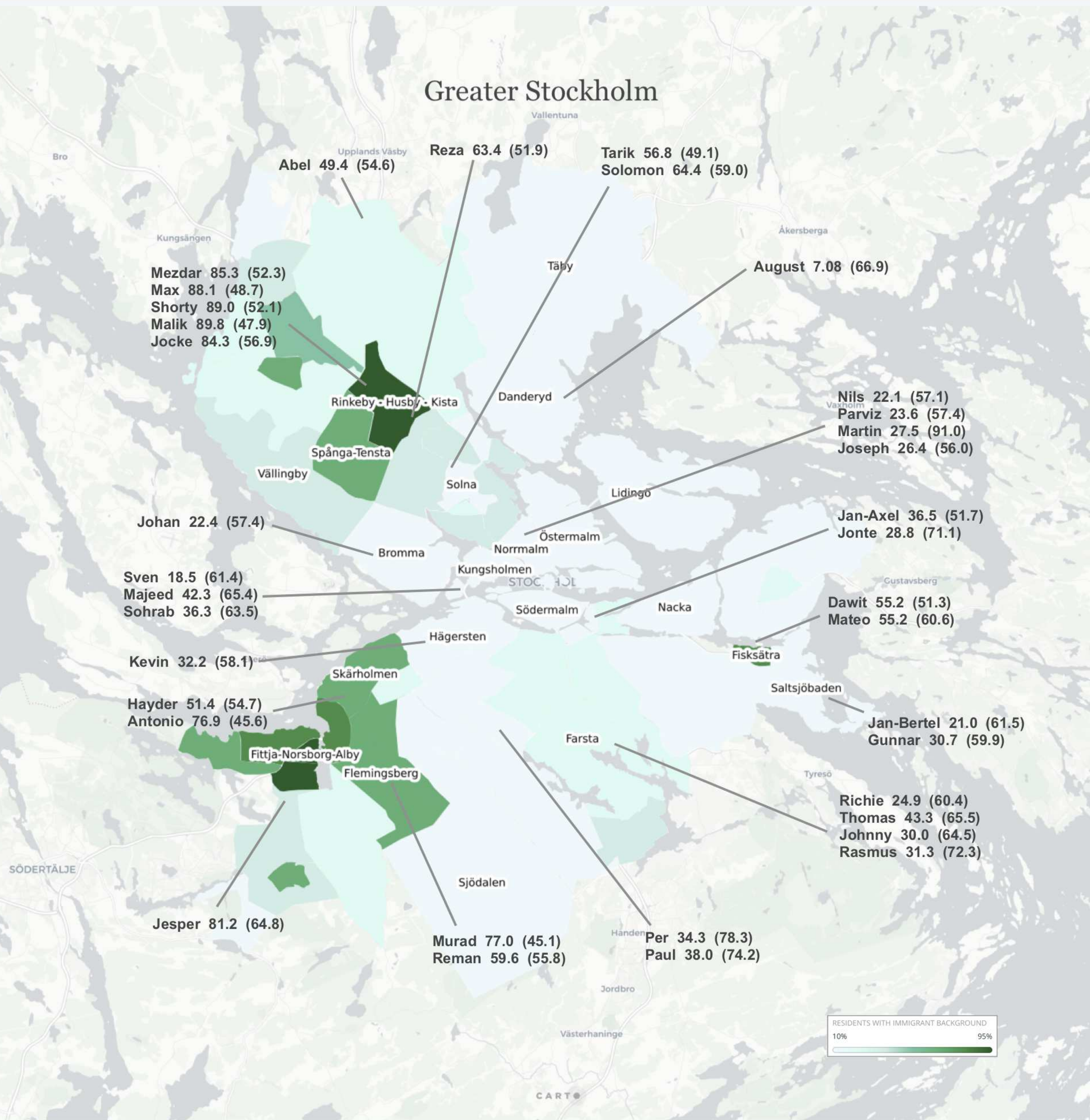


Figure 6.6: Diversity map of Stockholm, calculated from the percentage of residents of 'immigrant background' (foreign birth or with two parents of foreign birth; see legend in bottom-right corner) as reported 2010–2014 by Statistics Sweden. Participants are positioned as an overlay. They are listed by pseudonym, followed by lifetime neighborhood diversity index, followed by their nPVIV in CASUAL speech in parentheses.

Hayder

	Name	Network	Sex	Sexuality	Age	Ethnicity	Occupation	SEI	Language	Accent	Closeness
1	Contact_1	work	male	straight	45	Swedish	Building superintendant	21.8	Swedish	none	not close
2	Contact_2	work	male	straight	23	Bolivian	Building superintendant	21.8	Swedish	Spanish import	close
3	Contact_3	work	male	straight	26	Swedish	Building superintendant	21.8	Swedish	none	not close
4	Contact_4	work	male	straight	28	Greek	Security Guard	23.8	Swedish	none	not close
5	Contact_5	work	male	straight	28	Swedish	Security Guard	23.8	Swedish	none	not close
6	Contact_6	work	male	straight	23	Chilean	Security Guard	23.8	Swedish	none	not close
7	250 customers per week	work	50-50 fe/male	straight	20-30	mostly Swedish	-	55.0	Swedish	none	not close
8	Contact_8	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
9	*Reman*	gym	male	straight	24	Kurd	Social care / Security guard	24.4	Swedish	Hood	very close
10	Contact_10	gym	male	straight	35	Swedish	unknown	-	Swedish	none	close
11	Contact_11	gym	male	straight	28	Swedish	unknown	-	Swedish	none	close
12	Contact_12	gym	male	straight	28	Swedish	unknown	-	Swedish	none	close
13	Contact_13	gym	male	straight	23	Kurd	unknown	-	Swedish	Kurdish import	close
14	Contact_14	gym	male	straight	18	Lebanese	unknown	-	Swedish	Hood	close
15	Contact_15	gym	male	straight	21	Turk	unknown	-	Swedish	none	close
16	*Murad*	friend	male	straight	25	Egyptian	Manager furniture store	51.6	Swedish	Hood	very close
17	Contact_17	friend	male	straight	25	Serb	Sales clerk	29.5	Swedish	Hood	very close
18	Contact_18	friend	male	straight	25	Tunisian	Fire inspector	57.3	Swedish	Hood	very close
19	Contact_19	friend	male	straight	26	Serb	Internet technician	62.5	Swedish	Hood	very close
20	Contact_20	friend	male	straight	25	Serb	Social care	25.1	Swedish	Hood	close
21	Contact_21	friend	male	straight	25	Congolese	Sales clerk	29.5	Swedish	Hood	very close
22	Contact_22	friend	male	straight	25	Maroccan	Warehouse worker	16.4	Swedish	Hood	very close
23	Contact_23	friend	male	straight	25	Ghanian	Software developer	62.5	Swedish	none	very close
24	Contact_24	family	male	straight	58	Kurd	Home care	25.1	Kurdish	Import	very close
25	Contact_25	family	female	straight	55	Kurd	Assistant nurse	56.0	Kurdish	Import	very close
26	Contact_26	family	male	straight	23	Kurd	Social care	25.1	Kurdish / Swedish	Hood	very close
27	Contact_27	family	female	straight	18	Kurd	University student	-	Kurdish / Swedish	none	very close
28	Contact_28	family	male	straight	90	Kurd	retired	-	Kurdish	Import	very close
29	Contact_29	gym	female	straight	30	Kurd	Attorney	-	Swedish	none	not close
30	Contact_30	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
31	Contact_31	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
32	Contact_32	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
33	Contact_33	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
34	Contact_34	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
35	Contact_35	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
36	Contact_36	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
37	Contact_37	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
38	Contact_38	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close
39	Contact_39	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close
40	Contact_40	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close
41	Contact_41	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close
42	Contact_42	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close
43	Contact_43	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close

	1 Contact_1	*2* Contact_2	*3* Contact_3	*4* Contact_4	*5* Contact_5	*6* Contact_6	*7* 250 customers per week	*8* Contact_8	*9* *Reman*	*10* Contact_10	*11* Contact_11	*12* Contact_12	*13* Contact_13	*14* Contact_14	*15* Contact_15	*16* *Murad*	*17* Contact_17	*18* Contact_18	*19* Contact_19	*20* Contact_20	*21* Contact_21	*22* Contact_22	*23* Contact_23	*24* Contact_24	*25* Contact_25	*26* Contact_26	*27* Contact_27	*28* Contact_28	*29* Contact_29	*30* Contact_30	*31* Contact_31	*32* Contact_32	*33* Contact_33	*34* Contact_34	*35* Contact_35	*36* Contact_36	*37* Contact_37	*38* Contact_38	*39* Contact_39	*40* Contact_40	*41* Contact_41	*42* Contact_42	*43* Contact_43
1 Contact_1		1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 Contact_2			1	0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 Contact_3				0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 Contact_4					1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5 Contact_5						1	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6 Contact_6							1	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7 250 customers per week								1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8 Contact_8									0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9 *Reman*										1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10 Contact_10											1	1	1	1	1	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11 Contact_11												1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12 Contact_12													1	1	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13 Contact_13														1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14 Contact_14															1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15 Contact_15																0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16 *Murad*																	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17 Contact_17																		1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18 Contact_18																			1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19 Contact_19																				1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20 Contact_20																					1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21 Contact_21																						1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22 Contact_22																							1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23 Contact_23																								0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24 Contact_24																									1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25 Contact_25																										1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26 Contact_26																											1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27 Contact_27																												1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28 Contact_28																													0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29 Contact_29																														0	0	0	0	0	0	0	0	0	0	0	0	0	0
30 Contact_30																															1	1	1	1	1	1	1	1	1	1	1	1	1
31 Contact_31																																1	1	1	1	1	1	1	1	1	1	1	1
32 Contact_32																																	1	1	1	1	1	1	1	1	1	1	1
33 Contact_33																																		1	1	1	1	1	1	1	1	1	1
34 Contact_34																																			1	1	1	1	1	1	1	1	1
35 Contact_35																																				1	1	1	1	1	1	1	1
36 Contact_36																																					1	1	1	1	1	1	1
37 Contact_37																																						1	1	1	1	1	1
38 Contact_38																																							1	1	1	1	1
39 Contact_39																																								1	1	1	1
40 Contact_40																																									1	1	1
41 Contact_41																																										1	1
42 Contact_42																																											1
43 Contact_43																																											

Network - invandrare working-class = 8/13 = 61.5%
Network - svensk working-class = 0/13 = 0 %
Network - speaks 'hood' = 10/18 = 55.6 %

Table 6.14: Social network nodes and links for participant Hayder (pseudonym). Contacts 2, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, and Reman ($n = 13$) are close contacts for whom Hayder knew their occupation. Contacts 2, 17, 20, 21, 22, 24, 26, and Reman ($n = 8$) are close contacts out of that group whose SEI is < 50 and are identified as *invandrare*. Contacts 8, 16, 17, 18, 19, 20, 21, 22, 26, and Reman ($n = 10$) are close contacts who 'speak hood' out of the total group of close contacts who speak Swedish ($n = 18$; close contacts who speak only Kurdish ($n = 3$) are not included in the denominator).

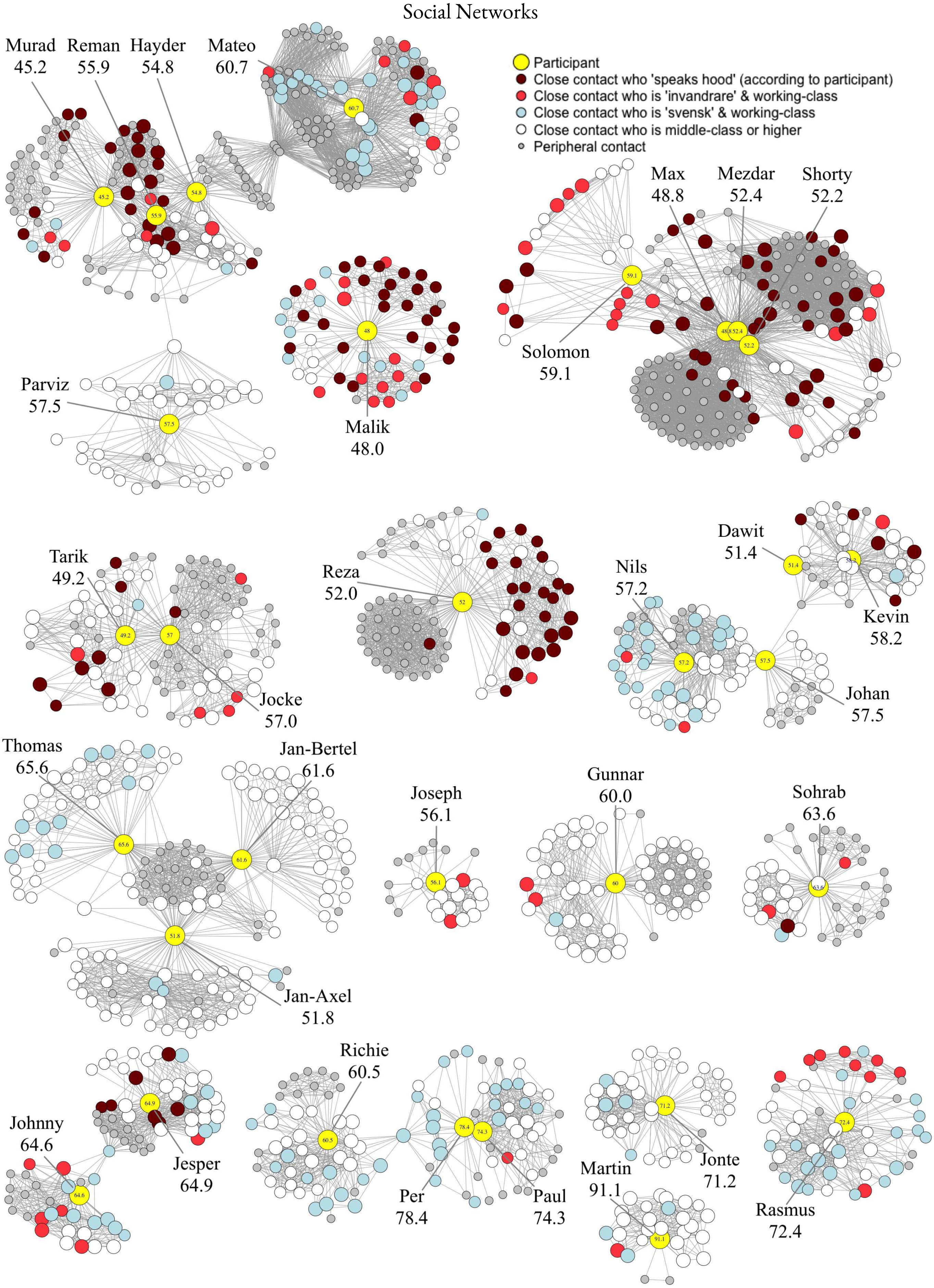


Figure 6.7: Social networks of the 29 speakers who participated in the social-network interview plus two speakers whose networks were partially mapped via participation in other ego stars (Dawit, Max). The 29 networks fall into 15 clusters. The node legend is located at the top. Network clusters are arranged in ascending order by nPVIV to the extent it is possible[a]. Low-nPVIV acquaintance clusters are generally found at the top, and high-nPVIV acquaintance clusters are generally found at the bottom. Participant nodes (yellow) are annotated with the participant's pseudonym followed by his nPVIV in CASUAL speech.

[a] As an example, Parviz's sister used to be Reman's sister's boss, forcing Parviz to appear higher than his close contacts would place him.

Pseud.	class	racialization	classmates_foreign_background	classmates_working_class	lifetime_neighborhood_diversity	network_svensk_wc	network_invandrare_wc	network_speaks_hood
Reman	0.0	*invandrare*	49	59	60	13	44	32
Malik	1.1	*invandrare*	57	81	90	39	59	51
Max	4.4	*invandrare*	85	73	88	-	-	-
Antonio	7.6	*invandrare*	53	60	77	-	-	-
Hayder	19.0	*invandrare*	37	38	51	0	58	56
Dawit	19.8	*invandrare*	14	28	55	-	-	-
Solomon	20.1	*invandrare*	49	66	64	6	83	54
Murad	21.7	*invandrare*	59	61	77	8	48	71
Abel	25.4	*invandrare*	33	64	49	-	-	-
Mateo	31.6	*invandrare*	45	47	55	71	23	5
Reza	34.8	*invandrare*	49	46	63	3	44	74
Mezdar	37.5	*invandrare*	73	75	85	13	70	85
Jocke	42.2	*invandrare*	31	48	84	10	38	55
Sohrab	48.6	*invandrare*	21	66	36	7	14	7
Shorty	51.2	*invandrare*	57	59	89	3	39	68
Kevin	64.0	*invandrare*	37	56	32	6	17	9
Tarik	73.4	*invandrare*	34	38	57	0	14	27
Majeed	76.2	*invandrare*	16	38	42	-	-	-
Parviz	91.1	*invandrare*	21	28	24	4	0	0

Pseud.	class	racialization	classmates_foreign_background	classmates_working_class	lifetime_neighborhood_diversity	network_svensk_wc	network_invandrare_wc	network_speaks_hood
Per	14.7	*svensk*	23	48	34	92	0	0
Rasmus	23.8	*svensk*	10	50	31	54	24	0
Johnny	26.9	*svensk*	24	71	30	50	38	0
Sven	34.7	*svensk*	32	55	19	-	-	-
Paul	38.2	*svensk*	27	76	38	53	7	0
Jesper	38.7	*svensk*	50	52	81	39	33	19
Richie	52.8	*svensk*	22	36	25	50	0	0
Thomas	60.8	*svensk*	28	59	43	27	0	0
Nils	64.4	*svensk*	15	24	22	56	5	0
August	73.5	*svensk*	13	30	7	-	-	-
Jonte	76.9	*svensk*	19	44	29	12	0	2
Martin	82.2	*svensk*	15	44	28	14	5	0
Gunnar	82.3	*svensk*	22	56	31	3	5	0
Joseph	84.1	*svensk*	30	35	26	0	17	0
Jan-Bertel	86.5	*svensk*	13	50	21	0	0	0
Johan	88.9	*svensk*	15	27	22	0	0	0
Jan-Axel	100.0	*svensk*	11	22	37	8	0	0

Table 6.15: Operationalizing six network properties: (1) the weighted average percentage of classmates of 'foreign descent' in grades 7–12 (classmates_foreign_background), (2) the weighted average percentage of working-class classmates in grades 7–12 (classmates_working_class), (3) the weighted average percentage of residents of 'foreign descent' for every neighborhood lived in (lifetime_neighborhood_diversity), (4) percentage of close working-class *svensk* contacts (network_svensk_wc), (5) percentage of close working-class *invandrare* contacts (network_invandrare_wc), and (6) percentage of close contacts who 'sound hood' when they speak (network_speaks_hood).

'Paki' ('blatte'), 'ghetto', or 'Rinkeby' were coded as *hood* in the dataset. If they gave the term 'brytning ('accent'), I asked them to clarify if they meant 'suburban accent' ('förörtsbrytning') or what is often referred to as 'import accent' ('importbrytning'), a colloquial term for late-onset L2-Swedish speakers. I chose to name this predictor 'Network - speaks hood' (network_speaks_hood) to reflect the fact that *hood* (*orten*) was the most common lay term for multiethnolect (see also Young 2018c:42).

To illustrate how the data was collected and catalogued for each participant, Table 6.14 shows an example of the data collected for participant Hayder. Hayder provided me with 43 names in his social network interview. To the left side of the table are the qualities of each contact, including the language he speaks with them.

Three columns are used for the calculation of network_svensk_wc: Column *SEI* (derived from column *Occupation*), *Ethnicity*, and *Closeness*. The number of close contacts who are listed as Swedish and have an SEI under 50 are tallied and number zero out of 13 close contacts for whom occupation was known (0%).

Three columns are used for the calculation of NETWORK_INVANDRARE_WC: Column *SEI* (derived from column *Occupation*), *Ethnicity*, and *Closeness*. The number of close contacts who are listed as not Swedish and have an SEI under 50 are tallied and number 8 out of 13 close contacts (61.5%).

Three columns are used for the calculation of NETWORK_SPEAKS_HOOD: Column *Language*, *Accent*, and *Closeness*. First, I tally the number of close contacts who use Swedish with Hayder. I then tally the number of them who are listed as 'Hood'. These number 10 out of 18 close contacts who speak Swedish (55.6%). Note two things with this calculation. First, the reason why the denominator is 18 instead of 13 is that the first two indices reflect missing information on occupation. Second, 18 constitutes the total number of speakers with whom Hayder uses Swedish. I do not include close contacts with whom he solely speaks Kurdish (contacts 24, 25, 28) when calculating the denominator because I want to build an index of what percentage of Swedish inputs (and, by further proxy, outputs) are characterized as 'Hood'.

The right side, indicated with 1s and 0s, constitutes the structure of his network. If a contact knows another contact, s/he is marked with a 1. If not, s/he is marked with a 0. This part will be incorporated into the analysis in Chapter 9 (p. 266).

Figure 6.7 contains a visualization of all of the actual social networks and their qualities. The three-way breakdown for close contacts[16] is color-coded. Close contacts who 'speak hood' are shown in dark red. Close contacts who are *invandrare* and working-class are shown in red. All the close contacts who 'speak hood' are *invandrare*, and most are also working-class (although some are middle-class, not shown in Figure). Close contacts who are *svensk* and working-class are shown in blue. I have rank-ordered from top to bottom and left to right the networks by nPVIV in CASUAL speech. The participants are shown in yellow, and their pseudonym and nPVIV in CASUAL speech are annotated beside each ego star. Twenty-nine speakers participated in the social-network interview. Two additional speakers, Dawit and Max, are included in the visualization because their names were mentioned in other speakers' ego stars, bringing the total number of names up to 31. There are, however, only 29 complete ego-star networks, and they fall into 15 clusters.

Table 6.15 assembles all the network calculations for each speaker into a single catalogue; proxy network calculations are on the left, and actual network calculations are on the right.

4.3 Analysis

In the illustration in Figure 6.7, the visual trend is that low nPVIV corresponds with more red and that high nPVIV corresponds with more blue and white. In the visualization of neighborhood in Figure 6.6, the general visual trend is that low nPVIV corresponds with darker-green spaces and high nPVIV with lighter-green spaces.

[16]The fourth category 'close contact who is middle-class or higher' is simply for the speakers who are neither *invandrare* working-class or *svensk* working-class.

			← *proxy networks*					→			← *actual networks*			→
	Model 1 *Internal only*		Model 9 *Classmates - Foreign background*		Model 10 *Classmates - Working-class*		Model 11 *Neighborhood Diversity*		Model 1′ *Internal only Subset (n=29)*		Model 12 *Network - Speaks 'hood'*		Model 13 *Network - Invandrare & svensk WC*	
Response variable	nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)		nPVIV (casual)	
Internal predictors		vif		vif		vif		vif		vif		vif		vif
(Intercept)	7.8 (4.7)$^{\circ}$		−5.1 (10.8)		−19.4 (11.6)$^{\circ}$		−1.0 (12.2)		11.0 (5.4)*		3.2 (13.0)		−7.1 (12.2)	
accent_1_long·*yes*	23.6 (1.0)***	1.4	23.6 (1.0)***	1.4	23.6 (1.0)***	1.4	23.6 (1.0)***	1.4	24.0 (1.2)***	1.4	24.0 (1.2)***	1.4	24.0 (1.2)***	1.4
accent_1_short·*yes*	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.6 (1.2)***	1.2	14.2 (1.3)***	1.2	14.2 (1.3)***	1.2	14.2 (1.3)***	1.2
accent_2_long·*yes*	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	18.3 (1.5)***	1.2	19.1 (1.7)***	1.2	19.1 (1.7)***	1.2	19.1 (1.7)***	1.2
accent_2_short·*yes*	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2	10.7 (1.2)***	1.2	10.8 (1.4)***	1.2	10.8 (1.4)***	1.2	10.8 (1.4)***	1.2
unstressed_long·*yes*	8.6 (1.8)***	1.0	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.0)***	1.8	8.6 (1.1)***	1.8	8.6 (1.1)***	1.8	8.6 (1.1)***	1.8
unstressed_short·*yes*	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.1)***	1.7	15.5 (1.2)***	1.7	15.5 (1.2)***	1.7	15.5 (1.2)***	1.7
coda_r·*yes*	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	7.9 (1.3)***	1.0	6.7 (1.6)***	1.0	6.7 (1.6)***	1.0	6.7 (1.6)***	1.0
phrase_final·*yes*	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0	6.9 (1.1)***	1.0	6.1 (1.3)***	1.0	6.1 (1.3)***	1.0	6.1 (1.3)***	1.0
log(lexical_freq)	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.3 (0.1)**	1.1	−0.5 (0.1)***	1.1	−0.5 (0.1)***	1.1	−0.5 (0.1)***	1.1
log(speech_rate)	6.0 (0.8)***	1.1	6.0 (0.8)***	1.1	6.0 (0.8)***	1.1	5.9 (0.8)***	1.1	5.8 (1.0)***	1.1	5.8 (1.0)***	1.1	5.8 (1.0)***	1.1
Social predictors		vif		vif		vif		vif		vif		vif		vif
age	-		0.6 (0.3)*	1.2	0.9 (0.3)**	1.0	0.5 (0.3)$^{\circ}$	1.4	-		0.4 (0.3)	1.4	0.5 (0.3)$^{\circ}$	1.2
classmates_foreign_background	-		−0.2 (0.1)*	1.2	-		-		-		-		-	
classmates_working_class	-		-		0.0 (0.1)	1.0	-		-		-		-	
lifetime_neighborhood_diversity	-		-		-		−0.1 (0.1)*	1.4	-		-		-	
network_speaks_hood	-		-		-		-		-		−0.2 (0.1)*	1.4	-	
network_invandrare_wc	-		-		-		-		-		-		−0.1 (0.1)	1.2
network_svensk_wc	-		-		-		-		-		-		0.1 (0.1)*	1.0
aic	218 859.3		218 853.1		218 857.3		218 853.8		171 587.9		171 585.3		171 588.4	
bic	218 970.6		218 980.3		218 984.5		218 981.0		171 695.7		171 708.5		171 719.4	
Log Likelihood	−109 415.6		−109 410.5		−109 412.6		−109 410.9		−85 779.9		−85 776.6		−85 777.2	
Num. obs.	21 000		21 000		21 000		21 000		16 394		16 394		16 394	
Num. groups: vowel	3 099		3 099		3 099		3 099		2 760		2 760		2 760	
Num. groups: speaker	36		36		36		36		29		29		29	
Var: vowel (Intercept)	70.1		70.1		70.1		70.1		75.8		75.7		75.6	
Var: speaker (Intercept)	96.3		66.3		76.4		67.1		100.4		72.1		68.7	
Var: Residual	1 906.1		1 906.1		1 906.1		1 906.1		1 989.1		1 989.1		1 989.2	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Table 6.16: Seven separate mixed-effects linear regression models. Models 1 and 9–11 contain the internal predictors of the full 36-speaker dataset. Models 1′, 12, and 13 contains contain the internal predictors for the 29-speaker subset. Models 9–13 all contain age as a predictor along with a social network measurement as a second predictor: Model 9 - *Classmates of foreign background*, Model 10 - *Classmates from working-class families*, Model 11 - *Lifetime neighborhood diversity*, Model 12 - *Close contacts in network who speak 'hood'*, Model 13 - *Close contacts in network who are invandrare working-class* and *Close contacts in network who are svensk working-class*. For categorical predictors, the reference category is in italics. Coefficients are indicated in the center column, standard errors in the parentheses, and variance inflation factors (VIF) to the right.

To test these visual assessments, I ran five new mixed-effects linear regression models in *lme4* in *R* with the following call framework:

> nPVIV ~ ACCENT_1_LONG + ACCENT_1_SHORT + ACCENT_2_LONG + ACCENT_2_SHORT + UNSTRESSED_LONG + UNSTRESSED_SHORT + CODA_R + PHRASE_FINAL + LOG(LEXICAL_FREQ) + LOG(SPEECH_RATE) + AGE + <u>SOCIAL_NETWORK_MEASURE</u> + (1|SPEAKER) + (1|VOWEL)

The outputs are catalogued in Table 6.16. Model 1 consists of the internal-only model from Table 6.6. Model 9 has CLASSMATES_FOREIGN_BACKGROUND as the social network measure. Model 10 has CLASSMATES_WORKING_CLASS as the social network measure. Model 11 has LIFETIME_NEIGHBORHOOD_DIVERSITY as the social network measure. Model 1′ is a new internal-only model constructed from the dataset of 29 speakers[17]. Model 12 has NETWORK_SPEAKS_HOOD as the social network measure. Model 13 has both NETWORK_INVANDRARE_WC and NETWORK_SVENSK_WC as the social network measure. The latter two network measurements are the only truly orthogonal measurements, which is why they were combined into one model. This was not the case for the other network measurements, which necessitated separate models.

AGE The models in Figure 6.16 all show a pattern for age that much resembles the social-class models in Table 6.13. The coefficients range between 0.4 and 0.9, indicating that an additional year in age would correspond to between a 0.4 and 0.9 increase in nPVIV (*less staccato*).

CLASSMATES - FOREIGN BACKGROUND Model 9 in Figure 6.16 shows a pattern that much resembles the pattern for the social-class parameters in Table 6.13. For every additional percent of classmates of foreign background one had in school, one's nPVIV is predicted to reduce by 0.2 (*more staccato*). This is easily conceptualized if one recalls the case studies for social class, which also was a 100-point measure with a coefficient of −0.2 (p. 156). The difference between 20% and 80% for the *classmates - foreign background* predicts a nearly 20-point difference in nPVIV: higher (*less staccato*) for 20% and lower (*more staccato*) for 80%.

CLASSMATES - WORKING CLASS Model 10 shows no significance for this predictor, and the coefficient sits squarely at 0.0. In light of the earlier findings on class, the most plausible explanation for this is that a variable like 'working-class' is not sufficiently rich to capture the split dynamic that operates within Stockholm's working class. As has been argued elsewhere (Mulinari & Neergaard 2004; Young 2018c), there are really two working classes in Stockholm. A school can be heavily working-class and *svensk* versus heavily working-class and *invandrare*. It is therefore less surprising that this parameter rendered few meaningful predictions on rhythm.

[17] It is not meaningful to compare information like AIC or BIC between the actual-network models and Base Model 1, because the observation count is different (36 participants versus 29).

Lifetime neighborhood diversity Model 11 shows a significant effect for this predictor with a coefficient of −0.1, which indicates that it would predict a 0.1 decrease in nPVIV for every additional percentage point of foreign-background residents in a speaker's neighborhood. This trend is generally clear in Figure 6.6, particularly to the left of the map. On the right, however, the picture is more complicated. This complication explains the weaker coefficient of −0.1 (versus −0.2) and also explains the high error value of 0.1. Working-class speakers like Per, Paul, and Rasmus have nPVIV values in the 70s whereas their lifetime neighborhood diversity scores are in the 30s. Elite speakers like Jan-Bertel or August have nPVIV values in the 50s and 60s whereas their lifetime neighborhood diversity scores are under 30. As discussed on page 37, working-class *svensk* speakers are all from neighborhoods that are slowly transforming demographically. It is only the middle and upper classes who have been able to maintain ethnically-homogeneous enclaves. The effect of neighborhood diversity as a predictor, *per se*, is mostly carried by the *invandrare* working class, which means it may be capturing some other extraneous predictor.

Network - speaks 'hood' Model 12 shows the same coefficient (−0.2) for this predictor as for *classmates - foreign background*, *social class PC1* and *SEI*. For every additional percent of close contacts who speak 'hood', one's nPVIV is predicted to decrease by 0.2 (*more staccato*). This finding matches several qualitative observations I had about the speakers. First, participant Mateo fits the ethnic and class profile of a typical multiethnolectal speaker (see top left in Figure 6.7 and Table 6.15), but his nPVIV is 60.7 – much closer to that of mainstream *svensk* speakers. If one examines his social network, one can see he has predominantly *svensk*, working-class friends. An opposing case is that of Shorty (see top right in Figure 6.7 and Table 6.15). His social-class position implies that he should speak with a much higher nPVIV (*less staccato*); he has a Bachelor's degree from a prestigious research institution, and his father is a secondary school teacher of mathematics (see Table 6.10). But his social network is 68% *speaks 'hood'*, which offers additional explanatory power for his unusually low nPVIV of 52.2.

Network - *invandrare* working class /Network - *svensk* working class Model 13 shows significance for *Network - svensk working class* but not for *Network - invandrare working class*. The coefficient for the latter (0.1) resembles that of Table 6.13 in that working-class *svensk* affiliation of any sort seems to correspond with higher nPVIV (*less staccato*). Again, this model implies that it isn't just multiethnolect that may be staccato, Stockholm's indigenous working-class variety may be especially non-staccato.

4.4 Section summary

The results in the preceding analysis can be summarized in the form of four main points.

1. Age shows a mild effect and has strong significance in every model run. Younger speakers are predicted to have generally lower nPVIV values (*more staccato*) than older speak-

ers.

2. Network qualities that pertain to a 'foreign background' correlate with lower nPVIV values (*more staccato*).

3. Network qualities that pertain to speaking 'hood' correlate with lower nPVIV values (*more staccato*).

4. Network qualities that pertain to lower-class *svensk* contacts correlate with higher nPVIV values (*less staccato*).

5. Network qualities have qualitative explanatory power for individual speakers whose nPVIV values contradict the traditional social-class trend (e.g., Mateo and Shorty).

Returning to the original research question, *Does the speech of Stockholm's racialized working class have lower rhythmic alternation than other varieties in the city; i.e., is it 'staccato'?*, the evidence points to the affirmative. Further, *svensk* working-class identity seems to correspond with high-alternation speech rhythm that differs significantly from the rhythm of both the *invandrare* working class and the elites. Speech rhythm seems also to be changing in apparent time with lower alternation among younger speakers and higher alternation among older speakers.

5 Chapter summary

The evidence presented in this sections supports an affirmative answer to the Research Question: *Does the speech of Stockholm's racialized working class have lower rhythmic alternation than other varieties in the city; i.e., is it 'staccato'?* The analysis shows that racialized identities and lower social-class identities – according to almost all class parameters and for male speakers – will predict lower nPVIV (*more staccato*) in CASUAL speech. Rhythmic alternation is defined as the alternation of strong and weak elements, and it is measured with the normalized pairwise variability index of vowels (nPVIV) using the energy-f_0-integral (EFI) from each vowel. nPVIV correlates strongly with the interaction of social-class parameters and racialization in CASUAL speech.

The findings also indicate that the staccato trend is stronger for young speakers. The findings also show that having had classmates of a foreign background or having lived in neighborhoods with many residents of a foreign background corresponds to lower nPVIV (*more staccato*) in casual speech. The findings also indicate that the metapragmatic observation of 'speaking hood' corresponds with low nPVIV (*more staccato*) in terms of one's social network; a higher number of close contacts who 'speak hood' corresponds with a speaker having a lower nPVIV measurement.

An additional finding is that the opposite is predicted to be true for the lower-class *svensk* speakers. Having an 'indigenous' lower-class identity corresponds with the highest alternation in the city (*least staccato*). Further, the analysis shows that having close contacts in adulthood who are *svensk* and of a lower SEI predicts higher nPVIV in casual speech. These trends, however, for the *svensk* working class are attenuated by age in the same direction as the findings on multiethnolect. In other words, the same models that show competing rhythmic patterns between the two working classes show the same coefficients for age (with minimal error): Older means less staccato; younger means more staccato. The topic of age and change in apparent time will be the key focus of Chapter 9.

Finally, network qualities have explanatory power at the individual level. Some speakers who have an nPVIV that deviates from the traditional social-class trend (e.g., Mateo and Shorty) also have a disproportionate amount of close contacts from other social groups than their own.

Chapter 7

THE PHONOLOGY OF RHYTHM IN STOCKHOLM

Chapter 6 presented evidence that rhythm, defined by the normalized pairwise variability index of vowels (nPVIV) of the energy-f_0-integral (EFI), is socially stratified. To describe it in simple terms, nPVIV has a ternary stratification in Stockholm: (1) low alternation (staccato) is generally produced by the *invandrare* working class; (2) high alternation (not staccato) is generally produced by the *svensk* working class; (3) medium alternation (somewhere in between staccato and not staccato[1]) is generally produced by the higher class. This result emerged from a model that examined social class as the first principle component (PC1) of six class artifacts (income, SEI, education, parental SEI, parental education, and taste). The model was supported by separate models for each of those class artifacts, and it was supported by separate models that used social network qualities to designate racial and class affiliations (i.e., the racialized binning and SEI of an ego star's close contacts).

It is often the case in scientific investigations that addressing one research problem releases a cascade of new problems. Now that the question of staccato rhythm in the speech of

[1]NB that the term 'staccato' is used here as a *relative* designation, since no perception analysis has been conducted on the envelope of alternation for the term 'staccato'.

Stockholm's racialized working class has been addressed (at least in terms of prodution), new questions suddenly become more compelling:

DURATIONAL AND ARTICULATORY VARIATION A number of sources, reviewed in the next section, have proposed that the low-alternation rhythm in multiethnolects is caused by various durational and articulatory transformations (e.g., monophthongization) at the segmental level. Since the nPVIV calculation includes segmental duration, these proposals will be explored and tested with the Stockholm corpus.

f_0 VARIATION Because the calculation of nPVIV includes mean f_0, there is a possibility that the rhythmic metrics may be capturing a reduction in f_0 excursion or declination.

INTENSITY VARIATION Because the calculation of nPVIV includes mean intensity, there is a possibility that the rhythmic metrics may be capturing a reduction in intensity variation.

Only the first of these three phonological components will be investigated in this chapter. This is mainly because addressing all three became too broad for the length constraints of this dissertation. Of the three, durational and articulatory variation have been handled the most in the literature on rhythm variation, so I will prioritize contributing to this discussion here. In Chapter 9, I also discuss the transformations in f_0 and intensity in my examination of change in apparent time.

1 Introduction

Recalling vocalic quantity in Swedish

Recall from my review on pages 43–45 that Central Swedish, to which Stockholm Swedish belongs, has nine vowels that divide into short and long allophones. They number seventeen allophones in total[2], each of which have significant qualitative as well as the obvious quantitative differences. Table 7.1 provides an overview. According to Eklund and Traunmüller (1997), the standard variants of LETA, SÖT, LUS, SOT, and LÅS are highly diphthongal in Central Swedish, the standard variants of DIS and TYP somewhat diphthongal, and the standard variants of LAT and NÄT monophthongal (1997:11).

According to Kuronen and Leinonen (2001:137), short vowels in Central Swedish are on average 67% the duration of long vowels, varying between 58% and 78%. This was assessed on a spoken corpus, sample size unclear. Helgason et al. (2013:537) found that short vowels vary between 54 and 64% the duration of long vowels, assessed on the reading aloud of isolated words.

[2] Seventeen instead of 18 because short ⟨e⟩ and ⟨ä⟩ have merged.

		Long Allophone				Short Allophone			
ortho-graphy	phoneme	short-hand	actual	lexical set	SweFA-bet	short-hand	actual	lexical set	SweFA-bet
i	/i/	[iː]	[iː]	dis	II#	[ɪ]	[ɪ]	disk	IH#
y	/y/	[yː]	[yː]	typ	YY#	[ʏ]	[ʏ]	flytta	YH#
e	/e/	[eː]	[ɪ͡e]	leta	EE#	[ɛ]	[ɛ̞]	lett	EH#
ä	/ɛ/	[ɛː]	[ɛː] - [æː]	nät - lär	AE# - AEE#	[ɛ]	[ɛ̞] - [æ]	lett - särk	EH - AEH #
ö	/ø/	[øː]	[øː] - [œː]	söt - dör	OE# - OEE#	[ø]	[ø]	dörr	OEH#
u	/ʉ/	[ʉː]	[ʉ̟ː]	lus	UU#	[ɵ]	[ɵ]	ludd	UH#
o	/u/	[uː]	[uː]	sot	OO#	[ʊ]	[ʊ]	rott	OH#
å	/o/	[oː]	[oː]	lås	OA#	[ɔ]	[ɔ]	lott	OAH#
a	/ɑ/	[ɑː]	[ɑː]	lat	AA#	[a]	[a]	lass	AH#

Table 7.1: Central Swedish vowels according to long and short allophones. They are listed in the following order: orthography, phoneme, the typical shorthand method of writing the allophone, the typical actual pronunciation, the lexical set, its unique *Forced Alignment of Swedish* alphabet letter (SweFA-bet). Note that SweFA-bet designates short vowels with an H before its accent mark. Diphthongal long vowels are shaded in gray according to how diphthongal their standard variant is (*darker*: more diphthongal, *lighter*: less diphthongal, *white*: monophthongal).

Vocalic duration and rhythm in multiethnolects

Various proposals have been made about the connection between rhythm and durational variation. I test each of the suggestions below for the present data.

Kotsinas (1988a) proposed that the impression of staccato for Rinkeby Swedish was due to "a tendency for many speakers to reduce the difference between long and short syllables"[i] (my translation, 1988a:267).

Bodén (2007, 2010), made a similar proposal for Malmö, Sweden's third city. To identify the source of the staccato effect, she examined vocalic durations in a small random sample of her speaker corpus. She compared the duration of phonologically long vowels between the standard Malmö variety and the local multiethnolect (Rosengård Swedish) and the duration of phonologically short vowels between the two varieties. No significant difference was found. In other words, phonologically-long vowels did not shorten, and phonologically-short vowels did not lengthen with respect to the standard Malmö samples (2007:29).

Hansen and Pharao (2010) similarly proposed that Copenhagen's staccato-like multiethnolect was due to transformations of phonological quantity; i.e., short and long vowels. Twelve speakers of multiethnolect and 12 speakers of standard Copenhagen Danish participated in a map task that targeted 33 test words. The authors found that speakers of multiethnolect generally had "equal duration of long and short vowels before syllables containing a full vowel." (2010:93) and that this was generally "due to shortening of long vowels rather than lengthening of short vowels." (2010:91).

Torgersen and Szakay (2012) took a slightly different view for their study of London's multiethnolect (Multicultural London English, MLE). They used the nPVIV algorithm and demonstrated that MLE in Hackney has less intervocalic durational contrast than the white working-class varieties in Havering. They proposed, however, that this could be due to the monophthongization of certain diphthongal vowels.

> What may be part of the cause of changes in the rhythmic patterns is a change in the duration of particular segments. We have shown how FACE, and to some degree GOAT, have shorter durations in Hackney than in Havering. The non-Anglos are in the lead in having near-monophthongal variants and also have a more syllable-timed rhythm than the other speakers. (Torgersen & Szakay 2012:838)

Just as Torgersen and Szakay (2012) propose that the lower alternation in MLE is connected to multiethnolectal speakers using monophthongal vowels, the higher alternation found among the *svensk* working class could be due to their use of diphthongal variants. In *Södersnack*, several of the long vowels are more diphthongal than their standard Swedish counterparts. These vowels are discussed in the next section.

Diphthongs in 'white' *svensk* working-class speech

The findings on rhythm in Chapter 6 indicate that the speech of Stockholm's *svensk* working class has the highest rhythmic alternation of any male social group in the city. The stratification of nPVIV is provided in Figure 7.4.

Recall also from my review on pages 47–48 that the variants [ɪeː] for NÄT and [oː] for SOT are much more diphthongal than their Central Swedish counterparts (Eklund & Traunmüller 1997), which could leave Södersnack with higher intervocalic durational contrast than other varieties in the city. Refer to Figure 7.3 ahead on page 190 and locate NÄT, LETA, LAT, and LÅS in the vowel chart. If one was to replace all occurrences of NÄT with LETA and all occurrences of LAT with LÅS, one's speech would become more diphthongal. This could potentially translate into a higher nPVIV, but the connection needs to be tested.

The connection between Södersnack and higher rhythmic alternation has never been investigated. However, since Bodén (2007) and Torgersen and Szakay (2012) suggest that monophthongs and lower rhythmic alternation are connected, the connection between diphthongs and higher rhythmic alternation should also be investigated.

2 Research questions

A production study cannot prove a direct connection between the perception of staccato and any phonetic feature. This would have to be investigated by means of a matched guise test that manipulated various segments to see whether the perception of staccato changed. Since this study is focused on patterns in production, and it has defined staccato as low intervocalic alternation in production, it seeks to identify any single segmental changes that lie behind that alternation. With this in mind, the following research question guides this analysis:

> RQ: *Is the variation in rhythm caused by (a) specific change(s) in the vowel system of certain speakers?*

For the purpose of staying within a manageable scope, I have formulated four specific questions that are subordinate to the overarching question.

> RQ1: *Are phonologically long vowels shorter in Stockholm's multiethnolect (the speech of lower-class invandrare) than in other varieties?*
>
> RQ2: *Are phonologically short vowels longer in Stockholm's multiethnolect (the speech of lower-class invandrare) than in other varieties?*
>
> RQ3: *Are phonologically long vowels more diphthongal in Stockholm's Södersnack (the speech of lower-class svensk) than in other varieties?*
>
> RQ4: *Are phonologically long vowels more monophthongal in Stockholm's multiethnolect (the speech of lower-class invandrare) than in other varieties?*

3 RQ1 and RQ2 analysis

This section addresses RQ1 and RQ2. I demonstrate that phrase-internal vowel quantity is systematically changing, but not in the most straightforward way. Long vowels, particularly prominent ones, are shorter in the speech of lower-class *invandrare*. Short vowels are longer if they are unstressed; prominent short vowels show no significant change. These interactions between durational allophone and prominence do not all have the same intensity of effect; some have more than others. Furthermore, phrase-final placement results in a completely different stratified effect for many of the vowels; lower-class *invandrare* speech has arrestingly longer phrase-final vowels, regardless of their phonological category.

At the most elementary level, the central tendencies[3] of vowel duration should be compared between various speaker clusters to check whether durational variation is the latent cause behind nPVIV variation or not. Such an analysis should, however, not ignore the many potentially confounding internal factors such as phrase finality and postvocalic rhotics. It should also not ignore the difference between prominent and non-prominent vowels. For example, unstressed long vowels in fluent speech are naturally shorter in duration than stressed short vowels. The following are mean durations in milliseconds for the four prominence-quantity categories:

[3]In terms of representing central tendencies, *means* demand more cleaning than *medians*, because vowels that occur within hesitations (e.g., *"I thiiiiiiiink so"*) can have a dramatic skewing effect on the mean unless outliers are clipped. In the case of the Stockholm data, extreme outliers have been discarded for nPVIV, making it difficult to appropriately vet the arguments and counterarguments for means and medians. Each choice has its benefits at the expense of some downsides. For example, boxplots are helpful ways to visualize the data, but the visualizations are hard-coded in R to show medians and quartiles. This use of medians, however, is bad for smaller sample sizes, which is the case when the data is broken up by individual prominent vowels. In the case of small subsets of the data, means will provide a better representation of central tendencies (Hozo, Djulbegovic, & Hozo 2005). But then again, if the distribution of measurements is excessively skewed, a mean figure can be somewhat misleading. Because I have selected the Welch *t*-test for as my test of significance, and this is a test of means, I use means for the sake of consistency.

prominent long	114.1 *ms*
prominent short	90.2 *ms*
non-prominent long	66.7 *ms*
non-prominent short	56.2 *ms*

In this analysis, vowels within phrase-final syllables were analyzed separately, and prominent and non-prominent vowels were also analyzed separately. Postvocalic rhotics were removed from the dataset. The full analysis, therefore, examines the following eight categories of vowels:

PHRASE-INTERNAL VOWELS	PHRASE-FINAL VOWELS
prominent long	prominent long
prominent short	prominent short
non-prominent long	non-prominent long
non-prominent short	non-prominent short

The social clusters include the following four categories of speakers:

lower-class *invandrare* speakers
lower-class *svensk* speakers
higher-class *invandrare* speakers
higher-class *svensk* speakers

Lower class is defined as having a Social Class (PC1) < 50; *higher class* is defined as having a Social Class (PC1) > 50. (see Chapter 6's Section 3.3 for an explanation of the Principle Components analysis). It is important, however to note that 50 represents the midpoint of the first principle component of social class within this dataset and not the midpoint within society. It is entirely possible, for example, that speakers who have a first principle component of 70 correspond more closely with what is generally considered to be the middle-middle class in Stockholm. Therefore, note that the terms *lower* and *higher* are more about relative placement[4],

3.1 Phrase-internal vowels

Phrase-internal vowels constitute all vowels except for the ones that occur within a pre-pausal, that is to say, phrase-final syllable. Figure 7.1 provides boxplots of the median normalized duration for phrase-internal acoustically prominent long vowels, acoustically prominent short vowels, acoustically non-prominent long vowels, and acoustically non-prominent short vowels from the CASUAL data in the Stockholm corpus. The data are divided into four social categories: (1) lower-class *invandrare* speakers, (2) lower-class *svensk* speakers, (3) higher-class *invandrare* speakers, and (4) higher-class *svensk* speakers.

[4]This becomes all the more important to remember in Chapter 9 when the conditional inference tree analysis splits the *invandrare* group at 73.4 (p. 258).

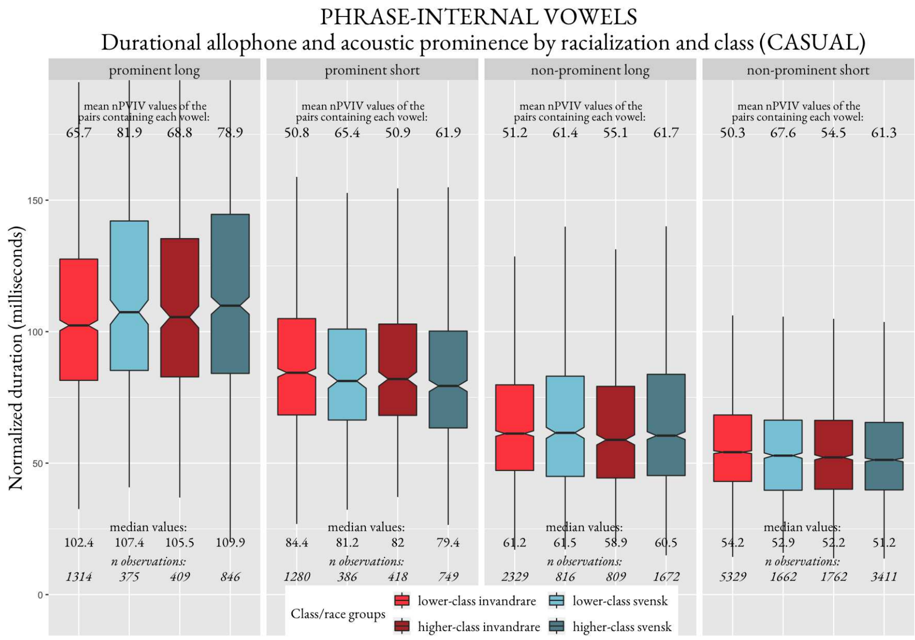


Figure 7.1: Box plots with 95% confidence intervals of normalized vowel duration according to durational allophone and acoustic prominence. Each allophone/prominence intersection is further divided up according to four social groups: lower-class *invandrare* speakers (left), lower-class *svensk* speakers (middle left), higher-class *invandrare* speakers (middle right), and higher-class *svensk* speakers (right). The actual median value and observation count for each boxplot is provided under each its respective boxplot. The mean nPVIV value of the pairs to which the vowel category belongs is itemized at the top above each its respective boxplot.

Vocalic durations were normalized by dividing each vocalic duration by the speaker's mean syllabic duration and then multiplying that result by the mean syllabic duration of the CASUAL dataset (148.3 ms).

FOR PHRASE-INTERNAL PROMINENT LONG VOWELS, Figure 7.1 shows that *invandrare* speakers have lower median durational values than *svensk* speakers overall. This is especially the case for lower-class *invandrare*. Welch two-sample *t*-tests[5] confirm that the durational differences between *invandrare* and *svensk* speakers are significant within and across social-class groups except for the difference between higher-class *invandrare* and lower-class *svensk*: *invandrare* produce generally shorter prominent long vowels, and *svensk* produce generally longer prominent long vowels. Social class, at least according to this approach to clustering, plays no significant role. This is summarized in the diagram below. Black arrows indicate significant differences; gray lines indicate uniformity. Each line is labeled with its *t*-test *p* value.

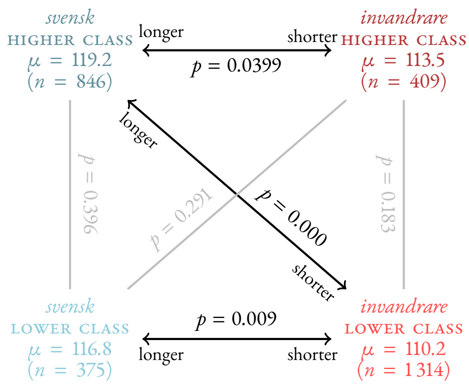
[5]Not all of the vowel measurements are normally distributed, which would typically motivate using the Mann-Whitney U test. However, since the means for each speaker group come from different sample sizes and have difference variances, the Welch *t*-test is the preferred test Ruxton (2006).

	Model 1	Model 2	Model 3	Model 4
	Prominent long	*Prominent short*	*Non-prominent long*	*Non-prominent short*
Response variable	⟵ *mean nPVIV of the two vowel pairs within which the vowel is situated* ⟶ ⟵ (*in casual speech*) ⟶			
(Intercept)	45.40 (2.90)***	42.22 (3.14)***	54.46 (2.16)***	67.02 (2.10)***
normalized duration (*ms*)	0.22 (0.02)***	0.19 (0.02)***	0.02 (0.017)	−0.17 (0.02)***
AIC	29 769.69	28 813.44	56 971.71	123 781.56
BIC	29 799.62	28 843.19	57 004.88	123 818.59
Log Likelihood	−14 879.84	−14 401.72	−28 480.85	−61 885.78
Num. obs.	2 944	2 833	5 626	12 164
Num. groups: vowel_stress	72	50	69	56
Num. groups: speaker	36	36	36	36
Var: vowel_stress (Intercept)	66.40	52.65	20.21	32.83
Var: speaker (Intercept)	95.71	145.85	75.13	80.41
Var: Residual	1 383.62	1 465.92	1 432.95	1 516.61

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Table 7.2: Statistical models with normalized duration as the predictor and nPVIV of the EFI as the response variable. Each model separately analyzes the durational effect of one of the four vocalic categories on nPVIV: (1) prominent long, (2) prominent short, (3) non-prominent long, (4) non-prominent short.

Prominent long vowels
in the casual speech of men
μ = 114.1 *ms* (n = 2 944)



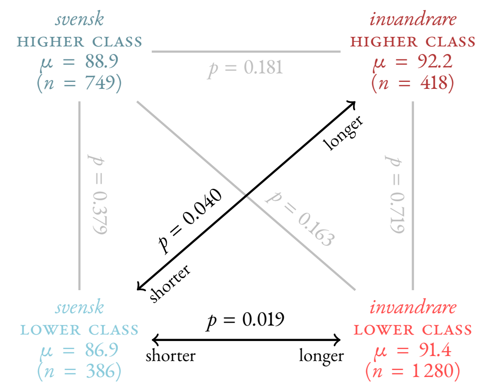
Given this trend, it would appear that the social stratification of prominent long-vowel duration is primarily a racialized stratification. *Invandrare* speakers have shorter prominent long vowels, and *svensk* speakers have longer prominent long vowels.

There also appears to be a straightforward relationship between the mean nPVIV values (of the pairs that contain these prominent long vowels) and the median duration of the vowels. The low-high-low-high pattern of 65.7 to 81.9 to 68.8 to 78.9 in Figure 7.1 resembles 102.4 to 107.4 to 105.5 to 109.9. To test this observation, I ran a mixed-effects regression analysis with mean nPVIV values of the pairs that contain prominent long vowels as the response variable and mean duration of prominent long vowels as the predictor. The results are shown in Table 7.2 Model 1. A single millisecond increase in the duration of a prominent long vowel is predicted to increase the nPVIV of the pairs it is situated in by 0.22.

It is important to note that the four models in Table 7.2 control for the effects of speech rate, because vocalic duration is normalized to the mean rate of the CASUAL subset of the corpus (148.3 *ms*).

FOR PHRASE-INTERNAL PROMINENT SHORT VOWELS, Figure 7.1 shows less of a difference between *invandrare* and *svensk* speakers. It does, however, appear to show variation across class lines. Welch two-sample *t*-tests confirm this partially. Lower-class *svensk* men have significantly shorter prominent short vowels than both lower-class and higher-class *invandrare*. This is summarized in the diagram below.

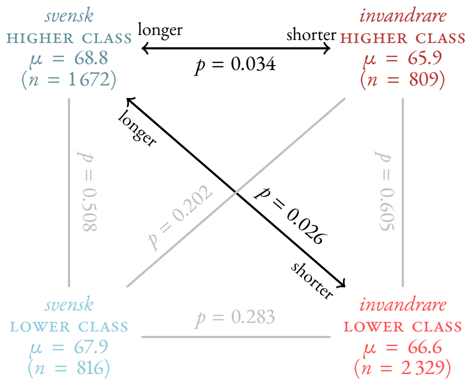
PROMINENT SHORT VOWELS
in the CASUAL speech of men
μ = 90.2 *ms* (n = 2 833)



Unlike the case for prominent long vowels, there does not appear to be a straightforward relationship here between the mean nPVIV values (of the pairs that contain these prominent long vowels) and the median duration of the vowels. In other words, the pattern in Figure 7.1 of 50.8 to 65.4 to 50.9 to 61.9 does not mirror 84.4 to 81.2 to 82.0 to 79.4. Nonetheless, in a mixed-effects regression analysis with mean nPVIV values of the pairs that contain prominent short vowels as the response variable and mean duration of prominent short vowels as the predictor, a significant effect was reported. The results are shown in Figure 7.2 Model 2. A single millisecond increase in the duration of a prominent short vowel is predicted to increase the nPVIV of the pairs it is situated in by 0.19.

FOR PHRASE-INTERNAL NON-PROMINENT LONG VOWELS, the trend mirrors that of prominent long vowels, but with fewer significant differences. According to the Welch two-sample *t*-tests below, the key difference is that higher-class *svensk* speakers produce longer non-prominent long vowels than both lower-class and higher-class *invandrare* speakers. None of the other differences are statistically significant, summarized in the diagram below.

Non-prominent long vowels
in the casual speech of men
$\mu = 66.7$ *ms* ($n = 5\,625$)

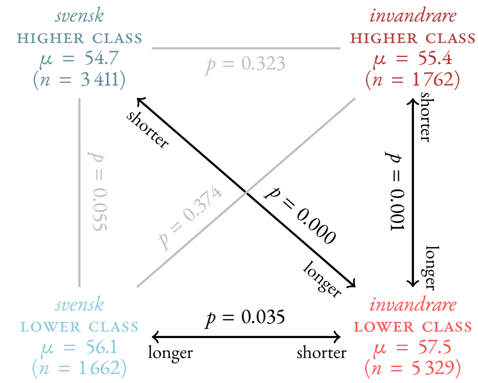


Given this trend, it would appear that the social stratification of non-prominent long-vowel duration is mostly a racialized separation that contains a mild class component.

Unlike the case for prominent long vowels, there does not appear to be a straightforward relationship between the mean nPVIV values (of the pairs that contain these prominent long vowels) and the median duration of the vowels. In other words, the pattern in Figure 7.1 of 51.2 to 61.4 to 55.1 to 61.7 only mildly mirrors the pattern 61.2 to 61.5 to 58.9 to 60.5. In a mixed-effects regression analysis, however, the mild visual resemblance is refuted. Table 7.2 Model 3 shows a non-significant effect (the coefficient of which is close to zero anyhow). It would appear that in the case of non-prominent long vowels, the stratification indicated in the above diagram is not responsible for staccato versus non-staccato rhythm. One possible explanation is that when a non-prominent long vowel is part of more staccato pair (lower nPVIV), its adjacent vowel is doing most of the adjusting, transforming into a size closer to it. Likewise, when a non-prominent long vowel is part of less staccato pair (higher nPVIV), a possibility is that this might typically be due to the adjacent vowel transforming into a size further from it.

For phrase-internal non-prominent short vowels, the stratification appears to only be related to the lower-class *invandrare*, who have a higher overall median duration than any of the other groups. Welch two-sample *t*-tests show that this is a significant trend. Lower-class *invandrare* have longer overall non-prominent short vowels than higher-class *invandrare*, lower-class *svensk*, and higher-class *svensk* speakers.

Non-prominent short vowels
in the casual speech of men
μ = 56.2 *ms* (n = 12 164)



As is the case for prominent short vowels, there does not appear to be a straightforward relationship here between the mean nPVIV values (of the pairs that contain these prominent long vowels) and the median duration of the vowels. In other words, the pattern of 50.3 to 67.6 to 54.5 to 61.3 does not resemble 54.2 to 52.9 to 52.2 to 51.2. However, a mixed-effects linear regression model refutes this visual trend, shown in Figure 7.2 Model 4. The model predicts that a single-millisecond increase in duration of non-prominent short vowels will reduce nPVIV by 0.17. This is a powerful result because the clear trend for rate, i.e. syllabic duration, is that it increases nPVIV. Despite constituting 12 164 of the 23 567 observations – i.e., 51.6% – an increase in duration for this vowel category is *not* predicted to increase nPVIV.

3.2 Phrase-final vowels

Phrase-finality partly reverses the trend seen for the phrase-internal vowels. When it comes to vowels in phrase-final syllables, it would appear that lower-class *invandrare* nearly always lengthen them beyond what is the norm for other speakers. Figure 7.2 shows the intersection of phonological length and prominence for racialization and class for phrase-final vowels. Most of the differences in the chart are statistically significant. But what stands out is that lower-class *invandrare* speakers have higher median lengths in all categories except for non-prominent long vowels (which constitute a relatively small percentage of phrase-final vowels overall). Prominence and phonological length appear not to be as relevant to this trend; rather, the phrase-finality seems to be the key determiner here.

A Welch t-test confirms this, summarized in the diagram below. Lower-class *invandrare* speakers lengthen the vowels within their phrase-final syllables significantly more than all other groups except for the *invandrare* higher-class speakers, whereby the difference is no longer significant. On the other hand, lower-class *svensk* speakers have the shortest phrase-final vowels, even shorter than those of higher-class *svensk* men, who in turn have longer phrase-final vowels than lower-class *invandrare* men.

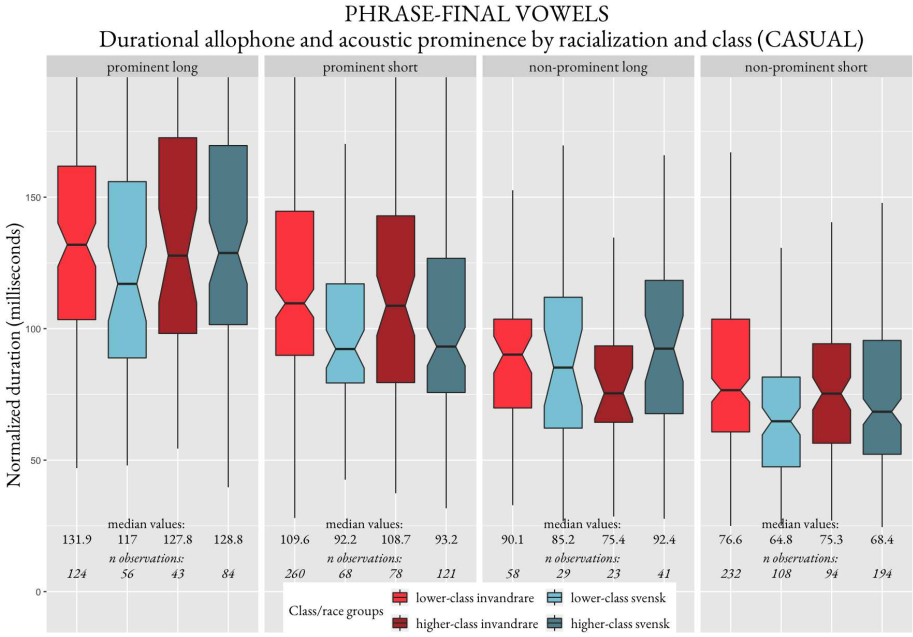
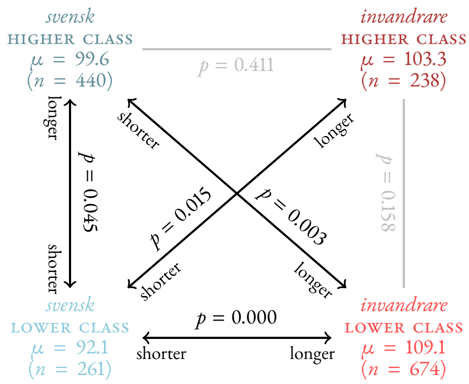


Figure 7.2: Box plots with 95% confidence intervals of normalized vowel duration according to durational allophone and acoustic prominence FOR VOWELS IN PHRASE-FINAL SYLLABLES. Each allophone/prominence intersection is further divided up according to four social groups: lower-class *invandrare* speakers (left), lower-class *svensk* speakers (middle left), higher-class *invandrare* speakers (middle right), and higher-class *svensk* speakers (right). The actual median value and observation count for each boxplot is provided under each its respective boxplot. The mean nPVIV value of the pairs to which the vowel category belongs is itemized at the top above each its respective boxplot.

VOWEL DURATION IN PHRASE-FINAL SYLLABLES
in the CASUAL speech of men
μ = 102.9 (n = 12 164)



4 RQ3 and RQ4 analysis

This section addresses Research questions 3 and 4:

> RQ3: *Are phonologically long vowels more diphthongal in Stockholm's Södersnack (the speech of lower-class svensk) than in other varieties?*

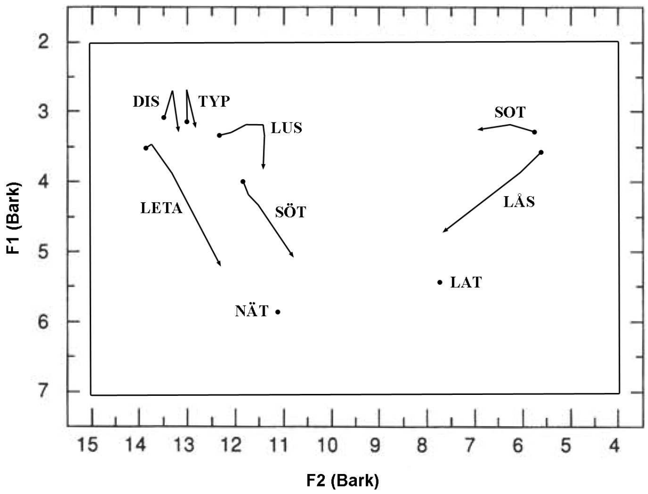


Figure 7.3: Long vowels in Central Swedish. Seven of the nine are considered to be diphthongs. (Eklund & Traunmüller 1997:11)

> RQ4: *Are phonologically long vowels more monophthongal in Stockholm's multiethnolect (the speech of lower-class invandrare) than in other varieties?*

Recall from Table 7.1 that seven of the nine Central Swedish long vowels are actually diphthongs, and refer to Figure 7.3 for their prototypical trajectory (Eklund & Traunmüller 1997). This analysis first compares the durations of each individual long vowel, long and non-prominent, across all speaker groups. It then examines F1-F2 trajectories for each vowel, long and prominent, across all speaker groups. It closes by measuring the vowel space area (VSA) for each speaker group in order to identify differences in vowel reduction.

4.1 Duration of individual prominent long vowels

Figure 7.5 shows boxplots for each vowel subdivided into the four social groups: lower-class *invandrare* speakers (left), lower-class *svensk* speakers (middle left), higher-class *invandrare* speakers (middle right), and higher-class *svensk* speakers (right). Figure 7.6 shows the mean duration for each vowel for each category along with the Welch *t*-test results for the difference between means between each social category. It shows that most of the vowels show no significant durational difference at all. Exceptions are lexical sets SÖT, LUS, LÅS, and LAT. Only LUS, LÅS, and LAT show significant shortening for the lower-class *invandrare*.

There is no significant durational difference for NÄT among the lower-class *svensk* speakers, which implies that the diphthongal *Stockholms e* (p. 47) is either not present or is not longer in duration than the standard variant. Higher-class *invandrare* speakers produce a shorter SÖT vowel than higher-class *svensk* speakers. Otherwise, SÖT shows no durational stratification. Higher-class *svensk* speakers produce a longer LUS vowel than the lower-class *svensk* and the lower-class *invandrare* speakers. Higher-class *svensk* speakers also produce a

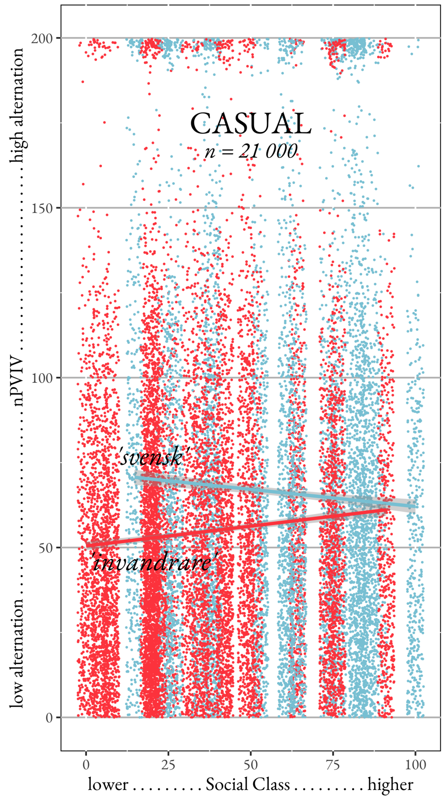


Figure 7.4: nPVIV alternation according to 'racialization' and class in the CASUAL vernacular of the current Stockholm dataset.

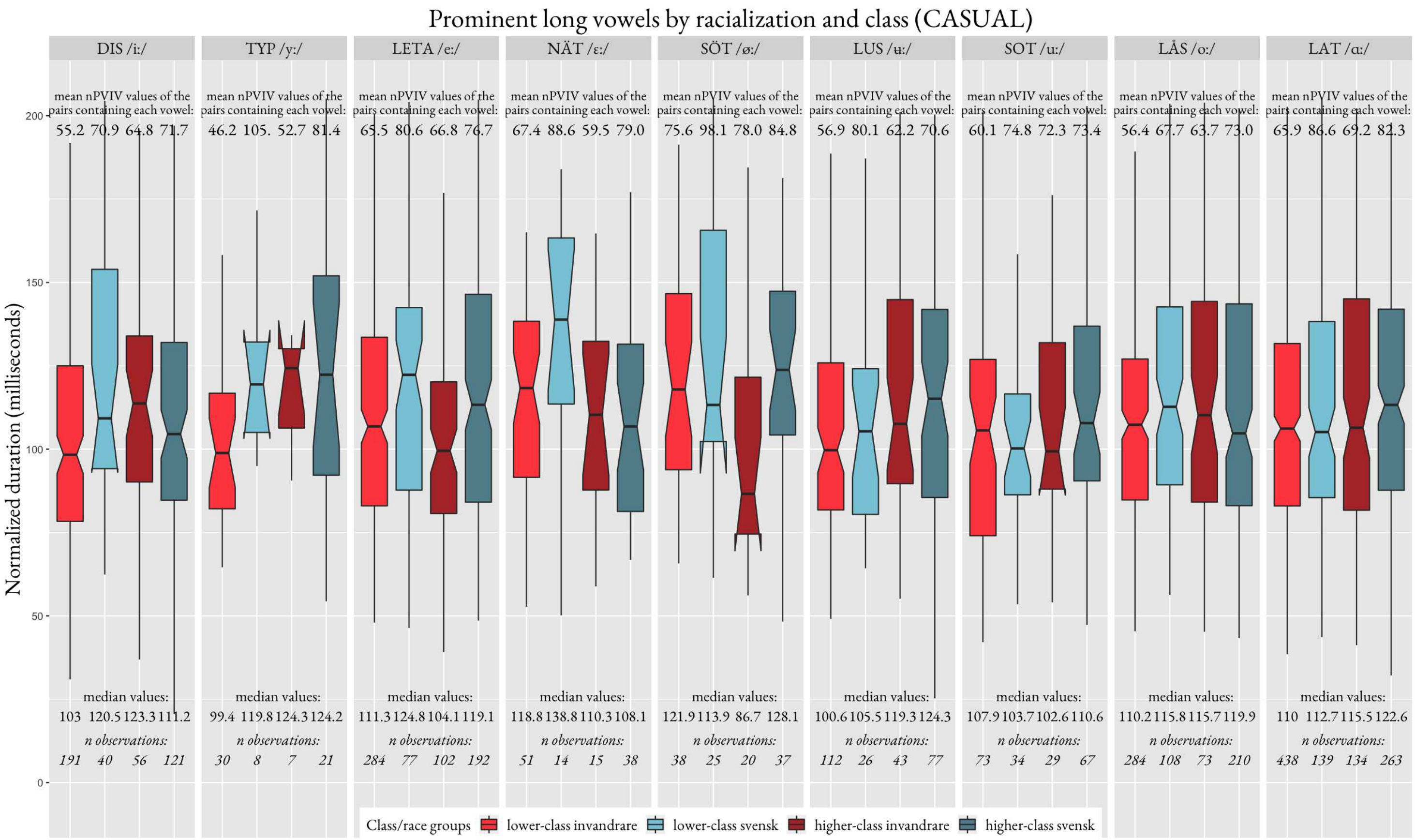


Figure 7.5: Prominent long vowels in the CASUAL dataset.

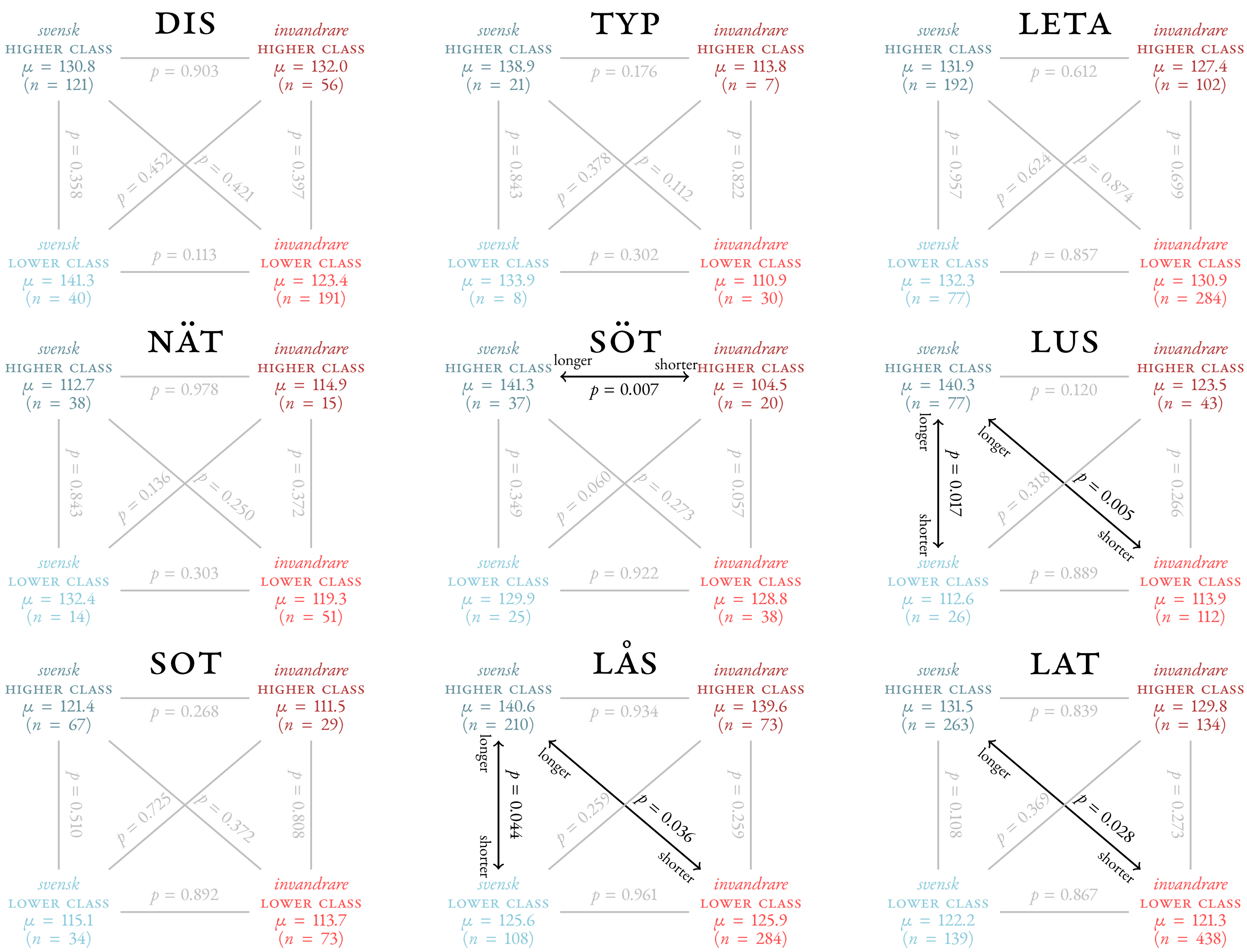


Figure 7.6: Median prominent long-vowel duration by class and racialization along with Welch two-sample t-test results for the difference between each social category.

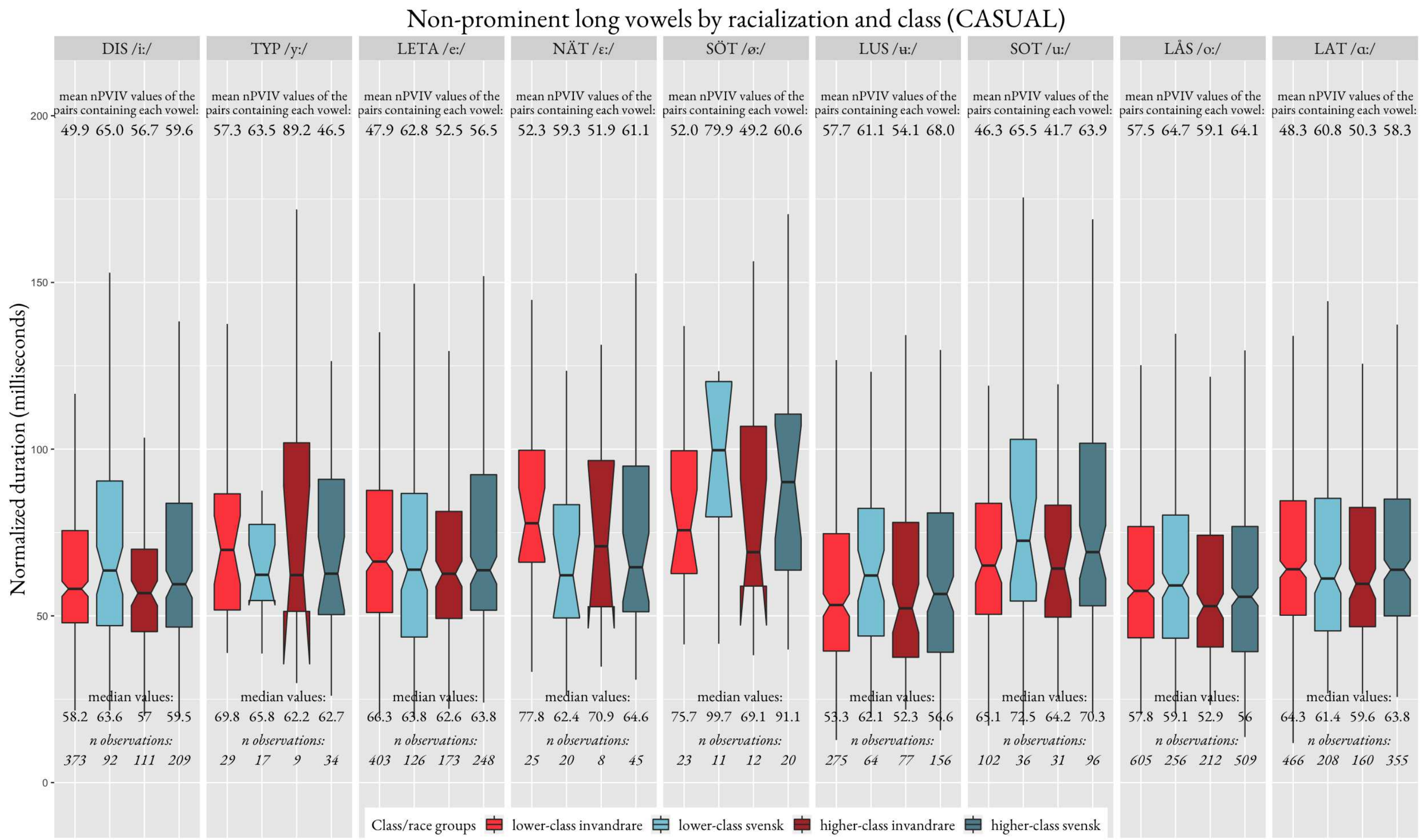


Figure 7.7: Non-prominent long vowels in the CASUAL dataset.

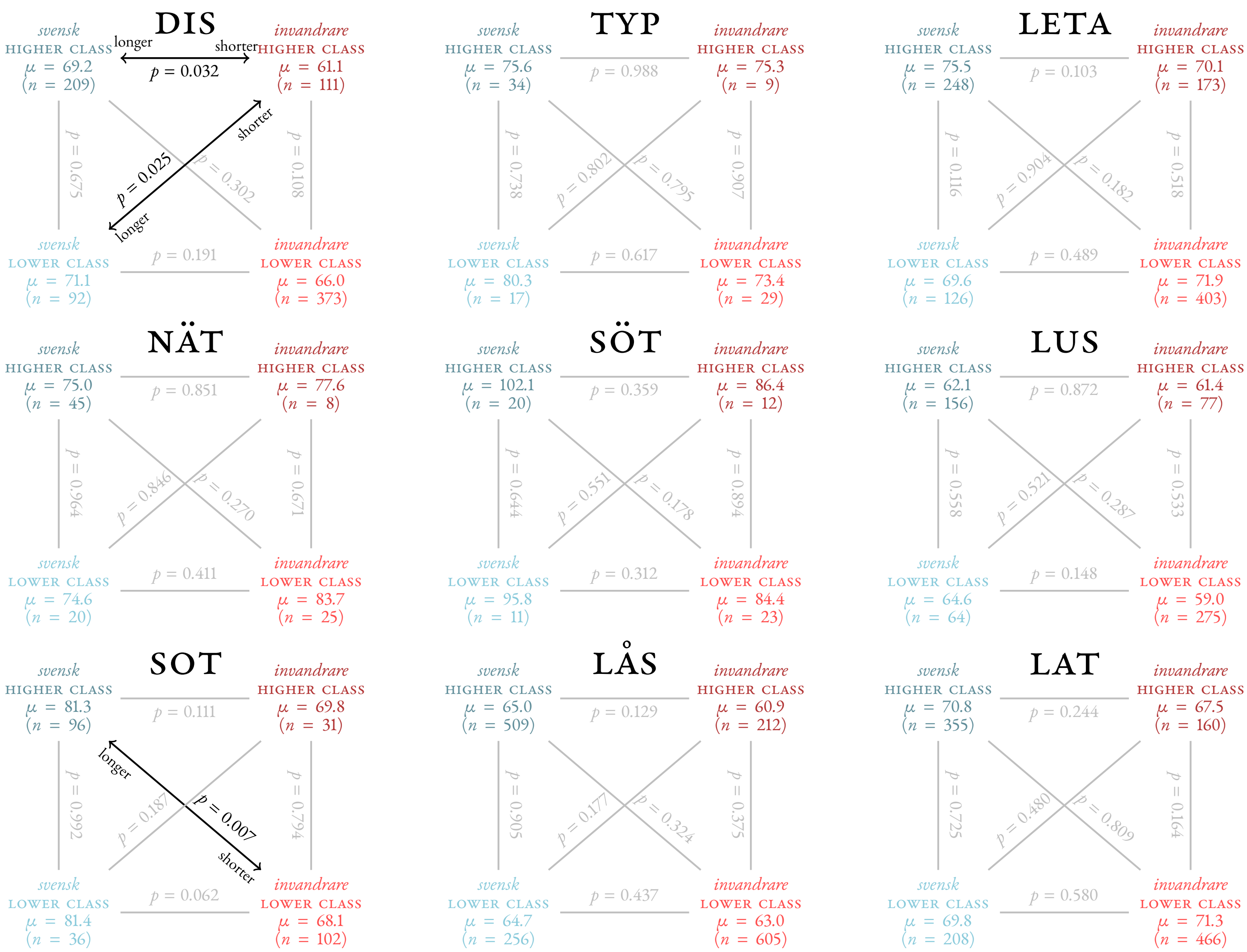


Figure 7.8: Median non-prominent long-vowel duration by class and racialization along with Welch two-sample t-test results for the difference between each social category.

longer LÅS vowel than lower-class *svensk* and *invandrare* speakers. Higher-class *svensk* speakers produce a longer LAT vowel than lower-class *invandrare* speakers.

Whereas there appeared to be an overall shortening of prominent long vowels among the lower-class *invandrare*, this analysis shows that only LUS, LÅS, and LAT appear to be key participants. In regards to staccato rhythm in multiethnolect, two possible interpretations are possible: (1) these three vowels are carrying the trend; (2) vowels within each set are shortening as appropriate to render the appropriate enregistered rhythmic pattern whereas other vowels in the same set are not.

4.2 Duration of individual non-prominent long vowels

Figure 7.7 shows boxplots for each non-prominent vowel subdivided into the four social groups: lower-class *invandrare* speakers (left), lower-class *svensk* speakers (middle left), higher-class *invandrare* speakers (middle right), and higher-class *svensk* speakers (right). Figure 7.8 shows the mean duration for each vowel for each category along with the Welch *t*-test results for the difference between means between each social category. It shows that even fewer of the non-prominent vowels show a significant durational difference than the prominent vowels. Only DIS and SOT show significant social differences in duration, and only SOT shows significant shortening for the lower-class *invandrare*.

Interestingly, DIS shows rather striking shortening among the higher class *invandrare* that warrants future investigation. As far as my own perceptions permit, there are no audible differences that I can report that correspond to this finding on duration.

As was the case for prominent vowels, there is no significant durational difference for non-prominent NÄT among the lower-class *svensk* speakers, which implies the diphthongal *Stockholms e* (p. 47) is either not present or is not longer in duration than the standard variant. It also implies that the LÅS-like LAT variant is not present or is not longer in duration than the standard variant.

Section 3 demonstrated that there was no clear shortening trend for non-prominent short vowels as a category for the lower-class *invandrare*. This analysis shows a similar result; yet, it also shows that the SOT lexical set is an exception. It is significantly shorter among lower-class *invandrare* than higher-class *svensk*. Its *p*-value is also barely non-significant in terms of its difference from the lower-class *svensk*. Given the fact that non-prominent SOT vowels (spoken by lower-class *invandrare*) only constitute 73 of the 21 000 vowels in the dataset, this would not be a major contributor to the staccato effect.

4.3 Variation in vowel-space movement

One would expect durational differences to capture at least some of the monophthongal transformations on long vowels. However, variation in vowel-space movement – such as the shortening of first and second-formant trajectories – can occur without shortening the actual duration of vowels. While it is not immediately clear how this would affect nPVIV, research questions 3 and 4 inquire about all aspects of diphthongization, including quality.

In order to calculate Euclidean distance, the first and second formant values (henceforth F_1 and F_2) were extracted from every percentage point within the temporal progression of every vowel between 5 and 95 percent. Measurements at 1 to 4 and 96 to 100 percent were not taken to avoid reading errors at the vowel boundaries. The values were extracted with a custom script in Praat (Boersma & Weenink 2017). They were then transferred to Excel and normalized according to Lobanov (1971):

$$F_i^{normalized} = \frac{(F_i - M_i)}{\sigma_i} \tag{7.1}$$

where F_i =	whichever formant measurement is to be normalized
where M_i =	the mean formant measurement for formant i (i.e., 1 or 2) of two exemplars of every vowel produced by the speaker at hand
where σ_i =	the standard deviation of formant measurements for formant i (i.e., 1 or 2) of two exemplars of every vowel produced by the speaker at hand

What the above formula does is it calculates a central 'base' F_1 for each speaker by taking the mean F_1 of two exemplars of each of the 17 Central Swedish vowels. It then calculates a central 'base' F_2 for each speaker by taking the mean F_2 of the same exemplars. The exemplars are taken from the reading task and are the same words for each speaker. After the base F_1 and base F_2 are calculated, the standard deviation is calculated for F_1 and F_2 in order to calculate the vowel 'spread' unique to each speaker. With these two measurements, every speaker's formant measurement can be adjusted according to the speaker's individual base and spread. Lobanov's (1971) technique is widely used in variationist sociolinguistics (Thomas 2011:160–171).

One challenge with Lobanov's (1971) technique is that it renders measurements that are meaningless if one were to compare them with findings from studies that have not normalized (or have normalized using a different technique). For example, if, at some later point, one wanted to compare these results with Eklund and Traunmüller's (1997) or Kuronen's (2000) vowel measurements, this would prove impossible with raw Lobanov-normalized scores. Therefore, I computed each normalized calculation back to the Hertz scale by multiplying back in the standard deviation from all 36 speakers and then adding the mean back in from all 36 speakers:

$$F_i^{normalized(Hz)} = F_i^{normalized} \cdot \sigma_i^{all} + M_i^{all} \tag{7.2}$$

where $F_i^{normalized}$ =	the normalized formant measurement i from Equation (7.1)
where M_i^{all} =	the mean formant measurement for formant i (i.e., 1 or 2) of two exemplars of every vowel produced by all 36 speakers in the dataset
where σ_i^{all} =	the standard deviation of formant measurements for formant i (i.e., 1 or 2) of two exemplars of every vowel produced by all 36 speakers in the dataset

Figure 7.9 shows the F_1-F_2 trajectories of the nine long vowels in prominent position according to each racialization/class subgroup. The trajectories are constructed by taking the median F_1 and F_2 values at every percentage point along the time progression of the vowel. The most immediately visible trend is a general lowering and fronting of back vowels among the *invandrare* speakers from all social classes This is represented by the red lines (light and dark) near SOT, LÅS, LAT, and SÖT. Another visible trend is the fronting of front vowels among the higher-class *invandrare* speakers, shown with the dark red lines near LETA and especially NÄT, which is remarkably fronted. These trends are beyond the scope of this investigation but are worth noting.

The vowel trajectories in Figure 7.9 would benefit from some sort of vertex decimation, but that is also beyond the scope of the current analysis. Despite their relatively messy trajectories, certain vowels show clear differences in distance covered between social categories. LETA is an example. For all speakers, the vowel begins close to the 1750-Hz mark for F_2 (x-axis). However, both *invandrare* groups end the trajectory at approximately 1600 Hz whereas both *svensk* groups end the trajectory at approximately 1500 Hz. To test this empirically, I calculated Euclidean distance between the 5% and 95% percentage points of the normalized values[6] and conducted Welch two-sample t-tests between each of the social categories. It is important to note that Euclidean distance was not calculated from percentage point to percentage point, because this would render misleading values. For example, examine the merged TYP/DIS variant produced by *invandrare* speakers, clustered between 2000 and 1800 Hz on the F_2 trajectory (x-axis) furthest to the left in the chart. The percentage by percentage extraction renders a treadmill-like trajectory, the actual distance of which is quite long. A linear measurement from 5% to 95%, however, removes the bias caused by this movement in place.

The results of the t-tests are modeled in Figure 7.10. They show a somewhat different pattern of variation than was evident in the durational analysis of prominent vowels in Figure 7.8. Where Figure 7.8 showed some durational variation for SÖT, LUS, LÅS, and LAT, Figure 7.10 shows qualitative variation for DIS, TYP, LETA, NÄT, LUS, and LÅS. The two variables they have in common, LUS and LÅS, do not necessarily validate each other. Lower-class *invandrare* variants of LUS are significantly shorter in duration than those produced by higher-class *svensk* speakers while also being *more diphthongal*. In terms of duration for LÅS both lower-class *svensk* and lower-class *invandrare* produce shorter variants than higher-class *svensk*. The analysis of quality, however, shows only a significant difference between higher-class *invandrare* (diphthongal) an lower-class *svensk* (monopthongal).

For the two highest vowels, DIS and TYP, the *invandrare* speakers have more or less merged them. This is something, I would add, that also is clearly audible in the dataset. What is not audible, in my opinion, is the result that lower-class *invandrare* produce more diphthongal

[6] One might argue that there is no need to calculate Euclidean distance on normalized values because the object of investigation is the hypotenuse of coverage rather than the x and y intercept of that coverage. Note, however, that the Lobanov technique of normalization also accounts for individual effects on *spread*. I am therefore of the opinion that Euclidean distance should be calculated on normalized measurements.

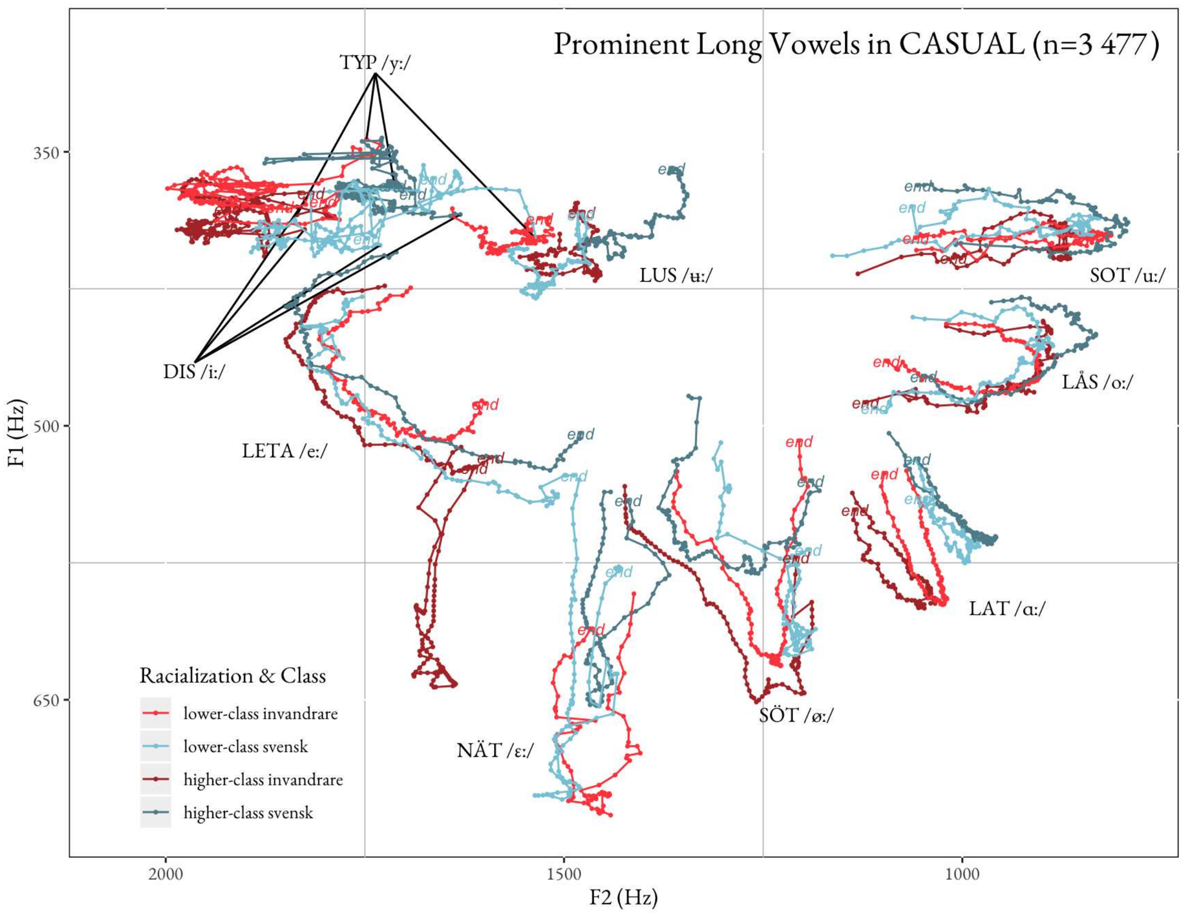


Figure 7.9: Lobanov-normalized F_1-F_2 trajectories of <u>prominent</u> long vowels in CASUAL dataset by class and racialization

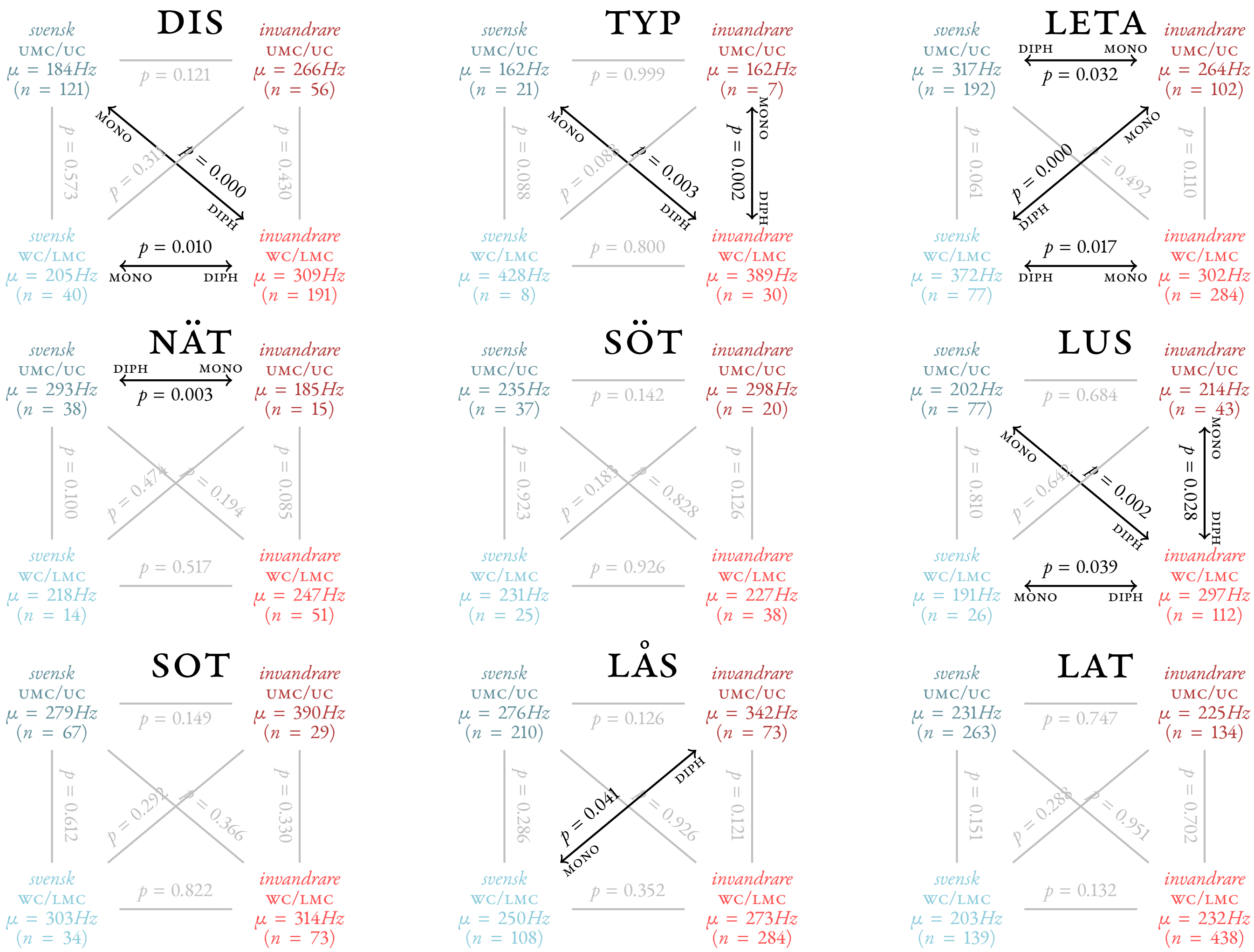


Figure 7.10: Mean F_1-F_2 Euclidean distance (from 5% to 95% of vowel progression) of <u>prominent</u> long vowels by class and racialization along with Welch two-sample t-test results for the difference between each social category.

DIS and TYP variants than the other groups.

LETA shows the clearest stratification between *invandrare* and *svensk* speakers. Regardless of social class, *invandrare* speakers produce significantly more monophthongal variants than *svensk* speakers. This is something that is also clearly audible in the dataset. Returning to the analysis on duration, however, one can see that this monophthongization does not match any significant durational shortening.

NÄT is significantly more monophthongal among higher-class *invandrare* than higher-class *svensk*. LUS is significantly more diphthongal among lower-class *invandrare* than any other social group. LÅS is more diphthongal among higher-class *invandrare* than lower-class *svensk*.

As it pertains to the typical diphthongs in *Södersnack* – i.e., pronouncing NÄT like LETA and LAT like LÅS – no clear trends are seen in terms of diphthongization. In the case of NÄT, the mean trajectory is shorter (218 Hz) than that of other higher-class *svensk* (293 Hz) and lower-class *invandrare* (247 Hz) and longer than that of higher-class *invandrare* (185 Hz). It is worth noting, however, that lower-class *svensk* speakers produce so few tokens of NÄT ($n = 14$) that the differences are unreliable. In the case of LAT, tokens abound, which indicates that if the backed/raised Södersnack-LÅS-like LAT variants are present, they are not necessarily accompanied by more diphthongization. Figure 7.9 does show, however, that both *svensk* groups do have a higher and more backed LAT than that of the *invandrare* groups, which implies a LÅS-like quality absent of diphthongization. Interestingly, looking at Figure 7.9, the distance between NÄT and LETA appears to be the smallest for the 15 tokens produced by higher-class *invandrare* speakers. It is possible that these tokens are associated with a particular stance, in which case a more qualitative analysis would be appropriate.

I conducted the same procedure for non-prominent long vowels. The results are shown in Figures 7.11 and 7.12. One may readily notice that the non-prominent vowel space appears much smaller with fewer clear separations between vowels. This is to be expected with unstressed vowels both because their full trajectories are not realized and because Swedish has quite a bit of vowel reduction. The latter point implies that most vowels will move toward a more central schwa position, causing them to overlap more.

Crucially, here, however is that the *invandrare* vowel space appears larger than that of the *svensk* speakers. For TYP, DIS, LETA, and LUS, red lines of both shades extend further to the left. For NÄT, SÖT, and LAT, red lines of both shades extend further downward. This is especially the case for lower-class *invandrare*, shown with the light red tracing; they reduce TYP, DIS, LUS, SÖT, and LAT less than any other social groups. Higher-class *invandrare* follow closely behind. Furthermore they seem to reduce LETA and NÄT less than any other social groups, including lower-class *invandrare*.

In contrast, lower-class *svensk* speakers appear to have the smallest vowel space of all social groups in their production of non-prominent long vowels. This is followed closely by higher-class *svensk* speakers.

In line with the visualization, the *t*-tests in Figure 7.12 also show that whenever *invandrare*

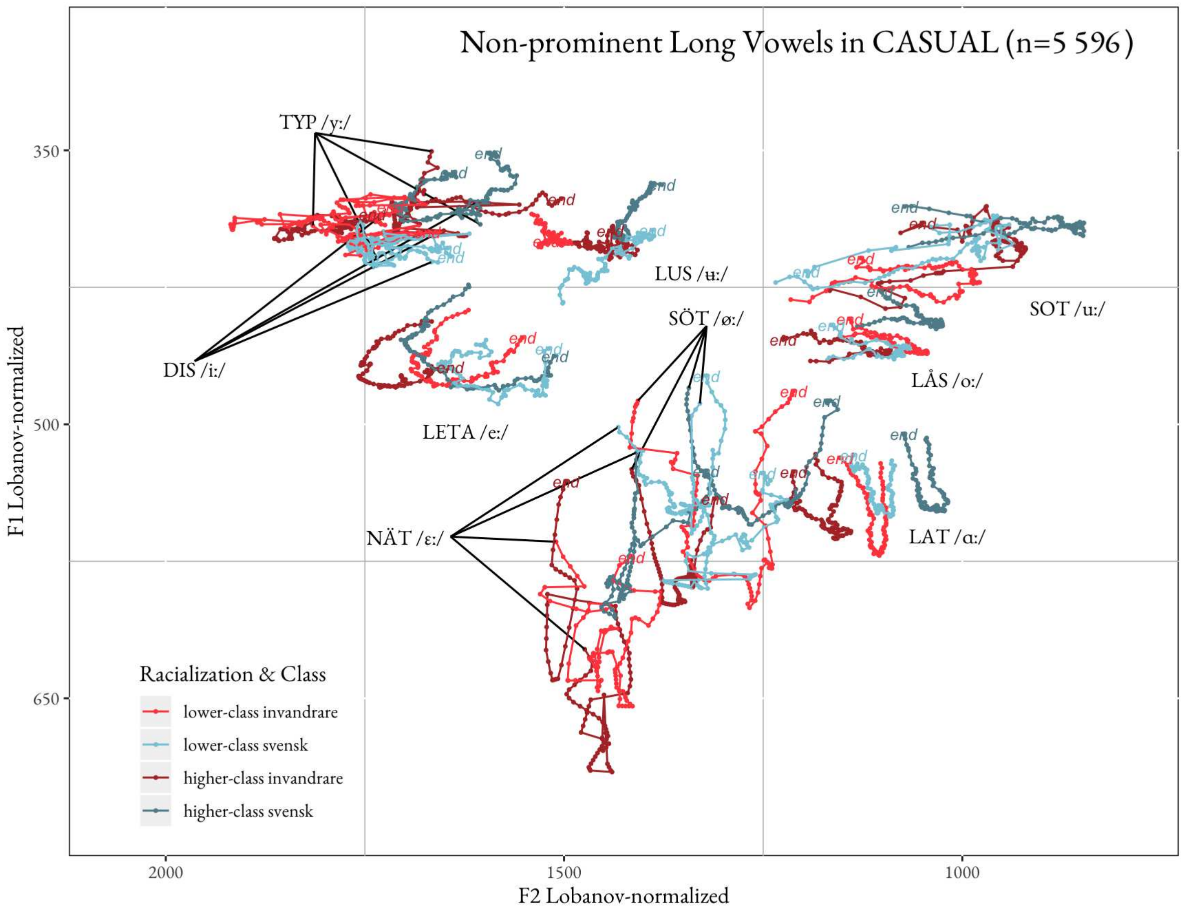


Figure 7.11: Lobanov-normalized F_1-F_2 trajectories of non-prominent long vowels in CASUAL dataset by class and racialization

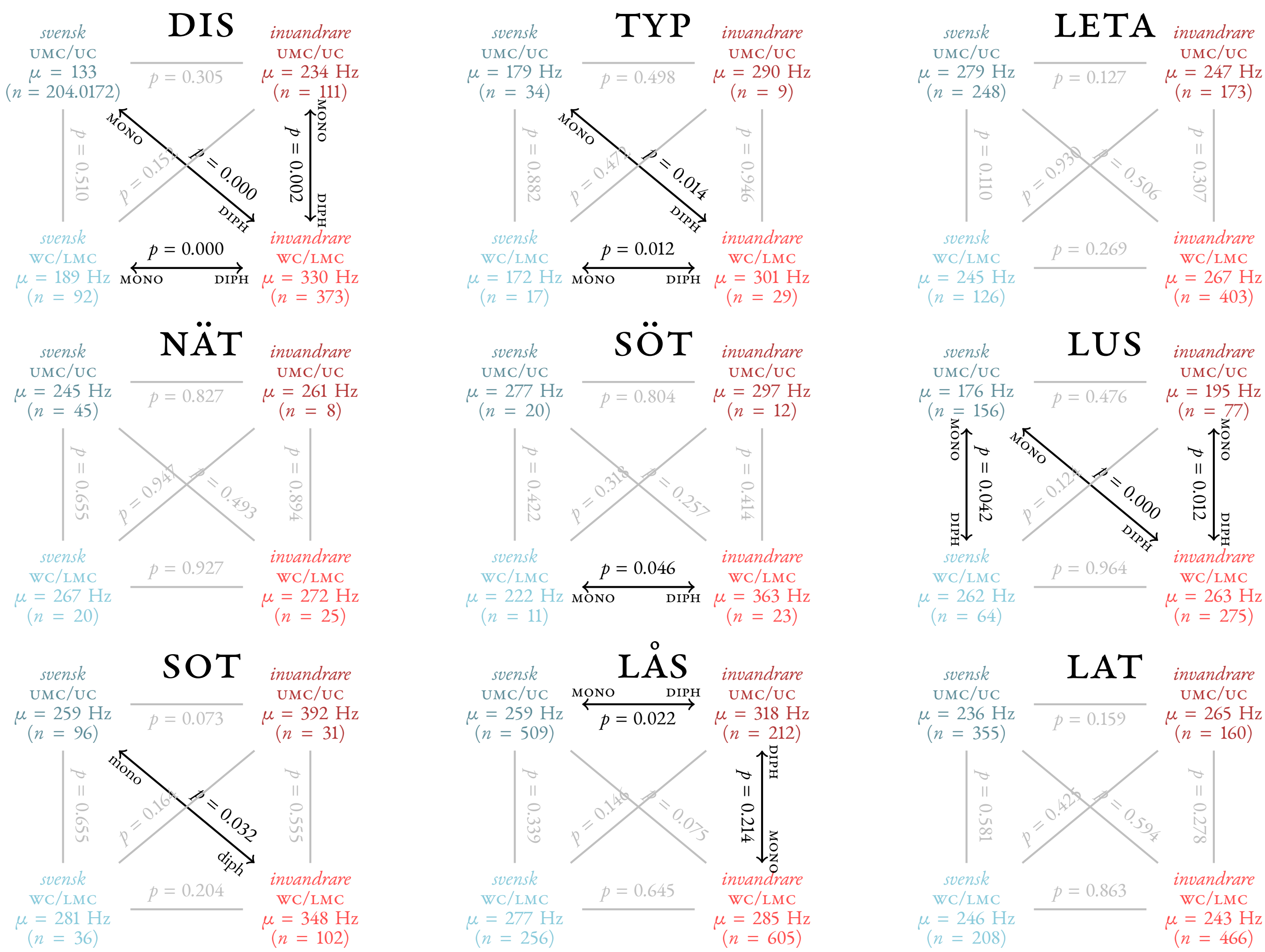


Figure 7.12: Mean F_1-F_2 Euclidean distance (from 5% to 95% of vowel progression) of non-prominent long vowels by class and racialization along with Welch two-sample t-test results for the difference between each social category.

speakers show a significantly-different trajectory, it is almost always longer than for those of the other social groups. This is the case for DIS, TYP, SÖT, LUS, and SOT, variants of which lower-class *invandrare* speakers produce more diphthongal variants than the other groups. In the case of LÅS, higher-class *invandrare* speakers produce the most diphthongal variant. What these results imply, taken together with the findings on prominent long vowels, is that no categorical monopthongization is underway. Rather, it appears that 'staccato' rhythm may be coercing some monophthongization in the case of prominent vowels and permitting fuller diphthongal realizations in the case of non-prominent vowels.

4.4 Reduction and the vowel space area (VSA)

Vowel space area (HZ^2) of male speech in Stockholm

	Prominent		Non-prominent	
	Long	Short	Long	Short
Working and lower middle-class *invandrare*	255 933	200 366	156 444	104 765
Upper-middle and higher-class *invandrare*	240 872	183 657	180 061	112 514
Working and lower middle-class *svensk*	236 825	180 896	112 504	93 407
Upper-middle and upper-class *svensk*	224 737	178 036	148 756	99 571

Table 7.3: Vowel-space area calculations (convex hull) for prominent and non-prominent long and short vowels for four groups of male speakers in Stockholm. The largest area within each vocalic category is underlined

The visual impression of the vowel space area (VSA), discussed above, implies that *invandrare* speakers reduce their long vowels less than *svensk* speakers in non-prominent production. To test this more thoroughly, I conducted two additional analyses. First, I mapped out median calculations for short vowels (since up until now, only long vowels have been examined). Second, I calculated the vowel space for prominent long, prominent short, non-prominent long, and non-prominent short vowels for each of the four speaker subgroups. This was done with the convex hull calculation described in Sandoval, Berisha, Utianski, Liss, and Spanias (2013:EL480). The results are shown in Table 7.3.

For all four categories, *invandrare* speakers have a larger vowel space. For prominent vowels, lower-class *invandrare* have the largest vowel space; for non-prominent vowels, higher-class *invandrare* have the largest vowel space. The main implication with this finding is that *invandrare* are leading in phonological change among men, which is implied by the clear lowering chain shift in Figure 7.9. Another implication with the findings in Table 7.3 is that *invandrare* are reducing their non-prominent vowels less than *svensk* speakers. This is also implied by the F_1-F_2 plot in Figure 7.11. *Invandrare* speakers produce most of these vowels much further away from schwa than *svensk* speakers.

5 Chapter summary

Duration of phrase-internal vowels

The results on phrase-internal vowels can be repackaged in the form of the durational difference between long and short vowels, which most resembles the wording of Kotsinas (1988a:267). These are provided in the schematic below and rank-ordered by magnitude of difference.

	Prominent Δ duration long − short	Non-prominent Δ duration long − short
lower-class *invandrare*	18.8 ms	9.1 ms
higher-class *invandrare*	21.3 ms	10.5 ms
lower-class *svensk*	29.9 ms	11.8 ms
higher-class *svensk*	30.3 ms	14.1 ms

Following Kuronen and Leinonen (2001) and Helgason et al. (2013), a useful way to conceptualize the above is to frame it in terms of percentage; i.e., short vowels have a certain durational percentage of long vowels.

	Prominent Δ duration short as a pct of long	Non-prominent Δ duration short as a pct of long
lower-class *invandrare*	82.9 %	86.3 %
higher-class *invandrare*	81.2 %	84.1 %
lower-class *svensk*	74.4 %	82.6 %
higher-class *svensk*	74.6 %	79.5 %

The above two schemas show that the differences between long and short vowels are shortest for lower-class *invandrare* speakers, followed by higher-class *invandrare* speakers, followed by lower-class *svensk* speakers, followed by higher-class *svensk* speakers. Re-examining the two research questions, it becomes clear that they are formulated too simplistically:

> RQ1: *Are phonologically long vowels shorter in Stockholm's multiethnolect (the speech of lower-class invandrare) than in other varieties?*
>
> RQ2: *Are phonologically short vowels longer in Stockholm's multiethnolect (the speech of lower-class invandrare) than in other varieties?*

The multiethnolectal patterns in nPVIV – that is, the nPVIV of the lower-class *invandrare* – appear to be a function of the shortening of prominent long vowels <u>and</u> the lengthening of non-prominent short vowels. The shortening of non-prominent long vowels, and the lengthening of prominent short vowels appear to play much less of a role as far as the social stratification of rhythm is concerned. The former is socially stratified but has no connection to nPVIV (compare matrix with Figure 7.2's Model 3); the latter has a connection to nPVIV but is only socially stratified between the lower-class *invandrare* and the lower-class *svensk* (compare matrix with Figure 7.2's Model 2). Another important point is that non-prominent short vowels constitute the most 'real estate' of any speech event. In the case of the Stockholm dataset, they make up more than half of all vowels. Therefore, I would argue that the lengthening of non-prominent short vowels must be contributing more to the staccato effect than the shortening of prominent long vowels. This conclusion, however, would need to be validated in a matched guise experiment where durations are manipulated and affective perceptions elicited.

The results resemble several accounts in the literature. Not only did Hansen and Pharao (2010) find that the difference between long and short vowels in Copenhagen multiethnolect had been eliminated, they identified the shortening of long vowels as the primary cause (2010:91). Kotsinas's (1988a)'s account appears also to be somewhat accurate in terms of framing what lies behind the staccato perception: "a tendency for many speakers to reduce the difference between long and short syllables" (1988a:267). The findings also offer an explanation why Bodén (2007) did not find any duration difference for short and long vowels when she compared Malmö multiethnolect with the received Malmö standard. As far as I am aware, her analysis did not code for prominence, something that appears here to be quite important.

Both research questions are affirmed partially. The shortening of prominent long vowels and the lengthening of non-prominent short vowels play a role in the staccato effect of multiethnolect. In as much as nPVIV is an adequate metric for rhythm, the shortening of non-prominent long vowels and the lengthening of prominent short vowels play much less of a role in the staccato production of lower-class *invandrare*.

Duration of phrase-final vowels

Phrase-final syllables complicate the picture by both superseding the shortening trend of long vowels while extending the lengthening of short vowels. In fact, the short/long, prominent/non-prominent distinction is less material here. Rather, the social distinction manifests itself in the difference between internal and phrase-final vowels. The amount by which the phrase-final vowel is lengthened, and the percentage by which it is lengthened is provided in the schema below.

Difference in duration between phrase-final and phrase-internal vowels

	increase ms	increase pct
lower-class *invandrare*	38.5 ms	54.6 %
higher-class *invandrare*	33.9 ms	48.8 %
higher-class *svensk*	29.4 ms	41.8 %
lower-class *svensk*	22.3 ms	31.9 %

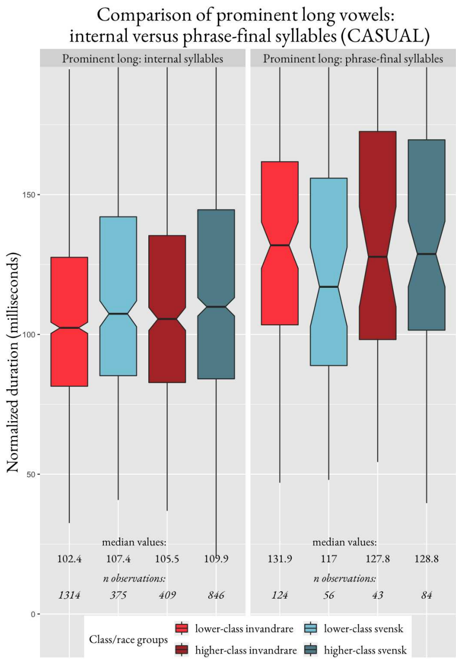


Figure 7.13: Comparing prominent long vowels in internal syllables with prominent long vowels in phrase-final syllables.

The distribution is striking. Lower-class *invandrare* lengthen phrase-final vowels by 16.2 milliseconds more than lower-class *svensk* men, 9.1 milliseconds more than high-class *svensk* men, and 4.6 milliseconds more than higher-class *invandrare* men. In fact, this trend is so antipodal to the phrase-internal trends that when the phrase-final and phrase-internal data are analyzed together, most statistical models are confounded and fail to capture any significant trends.

Therefore, returning to the question of staccato rhythm, variation in phrase-final lengthening is likely not contributing to lower alternation in multiethnolect. Rather, it appears to be its own separate variant.

The finding on phrase-final lengthening is particularly notable for prominent long vowels, because the pattern that seems to matter the most is that lower-class *invandrare* and *svensk* speakers are incongruent. Figure 7.13 shows the boxplots side-by-side. Higher-class *invandrare* and higher-class *svensk* speakers maintain the same general patterns for phrase-final syllables as for internal syllables. For lower-class speakers, however, the patterns are reversed. The *svensk* speakers maintain similar median durations whereas the *invandrare* speakers lengthen by over 50 milliseconds, maintaining a distinction from their *svensk* counterparts, albeit on the other side.

These results are in line with Bodén's (2010) discovery that speakers of multiethnolect in Malmö and Gothenburg have considerably more phrase-final lengthening than other speakers (2010:73–74). The results also resemble Kirkham's (2013) finding that Pakistani and Somali students in Sheffield produce longer ('tense'[7]) phrase-final /i/ than the standard regional 'lax' variant (2013:191–192). They also evoke Fagyal and Stewart's (2011) and Stewart's (2012) finding that the LHL boundary tone in Parisian is a racialized marker. Instead of lengthening the final vowel, this prosodic variant results in a lengthening of the penultimate vowel. And finally, they also evoke Britain's (1992) study that found young Maori speakers to produce high-rising terminal intonation contours at a higher rate than white New Zealanders.

Durational and articulatory variation of individual vowels

The results from the analysis on individual vowels can be summarized as follows:

1. *Invandrare* speakers from all social classes appear to be fronting DIS, TYP, and LUS, lowering and fronting NÄT, SÖT, and LAT, and lowering LÅS and SOT. Lower-class *invandrare* are the most advanced in fronting and raising of DIS, TYP, and LUS as well as the lowering of NÄT. Higher-class *invandrare* speakers are strikingly advanced in their fronting of NÄT and are also the most advanced in the raising and fronting of LETA as well as the lowering of SÖT, LAT, LÅS, and SOT.

2. *Invandrare* speakers, particularly those from the lower-classes, have the most distinctive prominent vowels in terms of F_1-F_2 trajectories. Prominent DIS, TYP, and LUS are more diphthongal. Prominent LETA is more monopthongal among lower and higher-class *invandrare* than among *svensk* speakers.

3. Only in the case of prominent LETA do lower-class *svensk* speakers show more diphthongal variants.

Re-examining the two research questions, it becomes clear that they too may have been formulated too simplistically:

> RQ3: *Are phonologically long vowels more diphthongal in Stockholm's Södersnack (the speech of lower-class svensk) than in other varieties?*
>
> RQ4: *Are phonologically long vowels more monophthongal in Stockholm's multiethnolect (the speech of lower-class invandrare) than in other varieties?*

The evidence points toward a refutation of RQ3 and RQ4. For RQ4, the picture is complex. Among prominent vowels, the high and front vowels (DIS, TYP, LUS) are more diphthongal for lower-class *invandrare* speakers; yet, LETA is more monophthongal. The most clear trend, however, is that non-prominent vowels are generally more diphthongal in the speech of lower-class *invandrare*, specifically, DIS, TYP, SÖT, LUS, and SOT. In this sense, the findings for nPVIV cannot be linked back to monophthongization as Torgersen and Szakay (2012) have proposed might be the case for MLE.

[7] Rather than the short/long distinction in Nordic languages, English vowels are categorized as 'tense' or 'lax'. The former corresponds approximately to 'long' and the latter to 'short' (see e.g., Jakobson, Fant, and Halle (1951:36))

Reduction and the vowel space area (VSA)

The analysis of the vowel space area VSA demonstrated that *invandrare* have a larger VSA for prominent long and short and non-prominent long and short vowels. For long vowels, they appear to be leading in phonological change, which is implied by the clear lowering chain shift in Figure 7.9, which is rendering the VSA larger. The findings also show that *invandrare* are reducing their non-prominent vowels less than *svensk* speakers, indicated in Table 7.3. This is also implied by the F_1-F_2 plot in Figure 7.11. There, it is apparent that *invandrare* speakers produce most of these vowels much further away from schwa than *svensk* speakers.

While perception is not the target of this study, the above findings implicate a fourth possible category to the staccato perception. The nPVIV metric used in this study conceptualizes the alternation of prominence as a manifestation of *duration*, f_0, and *intensity*. However, the salience of non-schwas where schwas are expected may also be contributing to the staccato effect. As I reviewed on page 57, this was implied by Kotsinas (1990) when she identified the full realization of the unstressed vowels in ˈ*va.rit* and ˈ*pla.ttor* as salient features in Rinkeby Swedish (1990:264).

Citations in original language

[i]"en tendens hos många talare att reducera skillnaden lång - kort stavelse" (Kotsinas 1988a:267)

Chapter 8

RHYTHM AND CONTEXTUAL STYLE IN STOCKHOLM

Chapter 7 found no one-to-one match between a single segmental transformation and the stratification seen for nPVIV. Rather, the analysis indicated that the rhythmic features are due to a coalescence of multiple segmental transformations. This coalescence manifests itself in the form of rhythmic variation or, in other words, the variation of intervocalic prominence contrasts. The finding lends weight to handling rhythm as its own variable.

The treatment of rhythm as a variable – in the classic variationist sense of the word (Weinreich et al. 1968:165–183) – permits me to turn to a more sophisticated question. Do speakers adjust their rhythmic patterns in more formal speech such that they target some hegemonic community norm? The ensuing analysis attempts to answer this by examining rhythm in two formal speech styles and comparing this to CASUAL speech. The first formal style, READING, is the recitation of a reading passage. The second 'hyper-formal' style, RADIO, is the recitation of the same reading passage whereby the speaker has been instructed to "sound like a radio announcer on *Radio Sweden*" (Section 3, p. 215).

The above concerns questions on change. Any shift by working-class speakers toward the intermediate rhythmic pattern of higher-class speakers implies that rhythm holds some degree of social salience for the former. This, in turn, can approximately indicate where staccato (and the *svensk* working-class' *non-staccato*) sits in its temporal evolution and where it may be headed. According to Labov (1972a, 2001), innovative variants that emerge into a community do so initially absent of detection. With time, some such innovations may become socially salient for a statistical majority of speakers, resulting in a stigmatization of the feature. Other innovations, like some variants in the Northern Cities Shift (Labov 2007; Preston 2010), remain undetected and eventually diffuse into the wider speech community.

In Chapter 5, Section 1.1, I documented a long history of observations about the staccato impression of multiethnolect (Bijvoet 2018; Bijvoet & Fraurud 2008, 2011, 2012, 2016; Bodén 2004, 2007, 2010; Fraurud 2003; Kotsinas 1988a, 1990, 1994b; Milani & Jonsson 2012; Nordenstam & Wallin 2002; Young 2014, 2018c). I have also found literature that insinuates some degree of salience among laypeople (Bijvoet 2018:115, Milani and Jonsson 2012:54, Young 2018c). The current chapter will therefore test whether the social stratification of rhythm also varies systematically across contextual styles.

1 Introduction

As I indicated above, any shift by working-class speakers toward the intermediate rhythmic pattern of higher-class speakers implies that rhythm holds some degree of social salience for the former. Intraspeaker variation – i.e., when a speaker adjusts the production of certain linguistic variants in certain speech domains – is one of the more challenging pursuits for variationist sociolinguistics. This is because the adjustment of certain socially-stratified variants is typically always accompanied by the *non*-adjustment of *other* variants. In other words, while many variants may occur only in the speech of lower social groups, only some of them are shed in more formal styles while others appear relatively immune to speech domain. It has been proposed by Labov (1972a) that this is a direct reflection of how salient certain variants are. As I come to discuss in Section 1.2, this is also a reflection of how long the variants have been in circulation, which means that stylistic adjustment can serve as a litmus test for the longevity of a variant[1]. In the upcoming Section 1.1, however, I will first focus on the notion of salience itself because salience is the key underlying mechanism behind style-shifting.

1.1 Salience

During the past ten years, the concept of *salience* has garnered renewed attention from linguists who have sought to unpack its meaning with the help of a social-constructivist framework. The term has traditionally been under-defined (Meyerhoff 2015:71), but my sense is

[1]It is not always the case that longevity will render a variant salient. In other words, salient variants are by necessity established, but not all established variants are salient.

that the earlier literature defines it as a nuanced extension of *noticeability*. Whereas noticeability implies conscious noticing, salience covers both unconscious *perceiving* and conscious *noticing* (Squires 2016). Breadmore, Krott, and Olson (2014) refer to the former as *implicit awareness* and the latter as *explicit awareness*. Drager and Kirtley (2016) similarly link this distinction to cognitive processes and borrow Lieberman's (2003) framework, referring to the former as *automatic* and the latter as *controlled*. Kerswill and Williams (2002:81) define salience as "in some way perceptually and cognitively prominent".

Trudgill (1986) argues that salience plays a role in whether certain linguistic features are accommodated (Giles & Phillip 1979) or not and points to the use of American features by British pop singers as an example. These singers generally limit themselves to five American features, namely, monophthongization of PRICE, addition of postvocalic /r/, using British BATH for American LOT, using British TRAP for American BATH, and inserting [d] in occurrences of intervocalic /t/ (1986:13). He argues that monophthongization of PRICE is targeted because the large phonetic distance between [aː] and [ɑɪ] renders it salient. The latter four are targeted because they involve the replacement, addition, or removal of entire phonemes in the British system, which also renders them salient. Trudgill goes on to observe that British speakers who move to the United States typically also follow this pattern of accommodation except for using British TRAP for American BATH. He proposes that this may be due to the fact that it is *too salient* in its American-ness (1986:18).

According to Trudgill's (1986) framework, salience is caused by properties that are both permanent and not so permanent. My interpretation from this framework is that these properties fall along a continuum. On one end of the continuum are the psychoacoustic properties such as the physio-articulatory characteristics of phones. On the other end of the continuum are socially constructed properties such as the social-indexical link between a variant and some group. In between these two poles, in the middle, lie the deeply constructed properties acquired within the Critical Period such as phoneme boundaries.

Kerswill and Williams (2002) point out, however, that Trudgill's explanation of salience is somewhat circular. This is particularly the case for the social-indexical side of the above continuum. That is to say, surely a variant must first be salient in order to have any indexical tie to any speaker group. Therefore, such a tie does not sufficiently explain salience (Kerswill & Williams 2002:89–90). They also argue that even the physio-articulatory criterion of phonetic distance and the middle criterion of phonemic distinction do not hold either. This is because their own data show Metropolitan-London ('Estuary English') vowel variants in the speech of youth in Reading and Hull, UK, that are phonetically distant from their conservative counterparts *and* constitute mergers of phonological contrasts. They propose, therefore, that extralinguistic cognitive, pragmatic, interactional, social psychological, and sociodemographic factors are more pivotal (2002:105).

In both Trudgill (1986) and Kerswill and Williams's (2002) discussions of salience, the focus lies on causes that operate behind salience rather than how we should define the notion salience. The assumption in both works – and, in my opinion, this is a reasonable assumption to make – is that the term means 'noticeable on some level'.

Labov (2001) does not discuss salience much and appears to define it as an issue of *variant priming* more than social indexicality. He uses *s*-dropping in Spanish as an example: the occurrence of the *-s* morpheme in an utterance has little to do with the functional demand for the plural morpheme (e.g., '*laS cosaØ bonitaØ*' instead of '*laØ cosaØ bonitaØ*') and more to do with whether *-s* was primed in earlier utterances or not (2001:558). Labov et al. (2011) frame salience in a similar way. They played a series of guises for listeners with differing distributions of *-in* for (*-ing*) and found that beyond 30%, the indexical strength of each additional *-in* variant reduces considerably. They conclude that speakers will react to small percentage differences at low frequencies, but not for the same percentage differences at higher frequencies (Labov et al. 2011:457).

Watson and Clark (2013) argue that

> linguistic units do not exist with salience 'attached' but rather that a particular linguistic feature can be more or less salient depending on both the local social context and the micro-linguistic context in which it appears. (Watson & Clark 2013:320)

Their argument is motivated by findings that show that listeners react differently to the same stigmatized variant depending on the previous utterance. Specifically, in the merger of SQUARE and NURSE, listeners did not perceive non-standard SQUARE when it was preceded by standard NURSE. Listeners were more likely to perceive non-standard NURSE when they the SQUARE-NURSE minimal pair occurred in a preceding utterance (2013:320).

Yaeger-Dror (1993) shares Labov (2001) and Watson and Clark's (2013) perspective. She conceptualizes salience as the end result of a scenario whereby the immediate situational ecology renders a segment cognitively accessible. She investigated the strategic use of variants of /r/ among 'Koiné' and 'Mizrahi' singers in Israel and coded each occurrence according to how phonetically and cognitively accessible the situational ecology rendered the variant (1993:203–204). For example, intervocalic position constituted phonetically salient occurrences (Lehiste 1970), and pre-pausal positions constituted non-salient occurrences (Levelt 1989). At first glance, singers showed seemingly nonsensical variability across contextual styles – using certain variants in their music, other variants in some interviews, and other variants in other interviews. However, after isolating phonetically salient occurrences, Yaeger-Dror found meaningful patterns in the variation.

Much in line with the Third Wave (Eckert 2012) of variationist analysis, Podesva (2011) takes a poststructural perspective and problematizes the notion of salience *as a property*.

> Taking such a perspective enables the analyst to treat salience not so much as a descriptive fact, but as a process. Forms are therefore not inherently salient, but made salient, however temporarily, by virtue of having been highlighted in discourse. (Podesva 2011:236)

He refers to Silverstein (2003) and Johnstone and Kiesling's (2008) work to make this point. According to Silverstein's (2003) account, the loss of the English second-person singular pronoun *thou* was facilitated by its overuse by Quakers. In 17th century England, *thou* was the informal second-person singular pronoun, and *you* was the second-person plural pronoun that also was used as the formal second-person formal singular. The Quakers rejected all formal pronouns on ideological grounds, which resulted in numerous memorable spectacles whereby the startling usage of informal *thou* in situations of social distance coincided with the presence of some 'cheeky' Quaker. This memorability, or salience, resulted in an eventual indexical linkage between *thou* and Quakers, a highly disfavored group, which drove *thou* out of usage by the greater community (Silverstein 2003:210–211).

Turning to Podesva's (2011) reference to Johnstone and Kiesling (2008), Pittsburghers who are unlikely to identify monophthongal /aw/ as local ('downtown' = 'dahntahn') in a perception experiment are *as likely* to use it in their own speech as they are the standard variant. In contrast, those who do identify it as local and stigmatized rarely use the variant in their own speech. When Johnstone and Kiesling (2008) examined individual speakers on a qualitative basis, they found that some older speakers were seemingly unable to tie any indexical meaning to phonetic variation. They argue that indexical associations are highly variable even within the same subgroup (such as working-class Pitsburghers). In their findings lies an implication that modern indexicalities may escape older speakers who have lost the sociocognitive plasticity to make new ideological connections between surface sounds and social categories (2008:23–26). This in itself exemplifies the need to conceptualize salience as highly variable and individuated.

Podesva (2011) promotes these examples to highlight the dynamism and the variability surrounding salience. Rather than being a property *per se*, salience is a fleeting characteristic that emerges in moments of distinction. He continues by pointing out two internal properties that increase the chances that a variant will emerge as situationally salient. The first is that the relative frequency of a feature will have an inverse relationship to its potential salience. In other words, the less it occurs, the more its occurrence has the potential to be distinct for listeners[2]. The second is surface-form acuteness; e.g., if one speaks with a fundamental frequency of 150 Hz and then migrates rapidly to 400, this will have a higher chance of registering as salient for listeners than a switch to 200 (2011:237).

Levon and Fox (2014) argue, however, that immediate or even recent contrasts should not be conflated with the embedded distinctions formed out of social experience. In their investigation of listener responses to *-in* for (*-ing*) and *f* for (*th*) in the UK, experimental priming had no significant effect on listener attitudes (except in the case of Northern listeners, which produced a very mild effect). The findings contradict Labov et al.'s (2011) results, outlined above, and Levon and Fox suggest that prior experiences with the variables in question can be so central that they override immediate situational effects (2014:209).

[2] This, however, according to Labov et al. (2011) is not a linear relationship. After a small number of occurrences, the slope flattens (2011:445).

Drager and Kirtley (2016) refer to such experiences – or rather, the mental generalizations of such experiences – as exemplars. In the context of linguistics, an exemplar is the mental representation of an event in which a linguistic form has occurred. Exemplars are registered moment by moment and filed in 'exemplar clouds' based on how similar they are to other clusters of exemplars (2016:4). While higher recency and higher frequency can make certain exemplars weigh more, a lifetime of exemplars counts toward this reckoning of recency and frequency. The result is, therefore, a balance sheet of competing 'exemplar weights' that take all of these factors into account (2016:13). Given sufficient exemplar weighting, the *production* of a specific variable can then also be influenced.

> If speech is stored in the mind as socially indexed and phonetically rich episodic memories, an individual's speech production and perception can be influenced by patterns present across the exemplars even when they don't notice the patterns. [...] It is possible, therefore, that production can be biased towards certain linguistic variants and that the bias can be socially motivated, while the speaker remains unaware of the association (Drager & Kirtley 2016:9).

The ensuing analysis does not directly investigate the current salience of rhythm in Stockholm. This would require the incorporation of a perception experiment. However, stylistic changes *in production* are directly connected to the perceivability of a variant's social meaning by the statistical majority of speakers. This connection between salience and production will be examined more closely in the following section.

1.2 Stylistic variation

This chapter attends to style rather than salience, per se. Nonetheless, salience is the pivotal underlying mechanism behind stylistic variation in variant production. A fundamental assumption in variationist sociolinguistics is that the degree to which a statistical majority of speakers adjust variants in formal speech reflects the longevity of the variant. This is because the adjustment reflects the salience of the variable at that point in time (for that statistical majority of speakers), and the salience of the variable is a reflection of how long the variant has been in circulation. In this section, I provide a review on the connection between social salience, intraspeaker variation, and change. Because of the findings I outline below, it has become standard practice to use intraspeaker variation as a proxy for gauging the chronological development of sound change.

In his linguistic investigations of New York and Philadelphia, Labov (1972a, 2001) discovered that some of the variants unique to working-class casual speech were produced at lower rates when those same speakers moved into more formal contextual styles. He also found that other variants were produced at the same rates, regardless of contextual style. This trend made it necessary to find a term that meant 'noticeable' without the strong connotations of awareness. As I reviewed above, *salience* satisfies this demand; the word indicates perceivability and remains agnostic about awareness.

The very same speakers who did not adjust some variants did adjust others, so the idea that they simply *chose* not to participate in style-shifting was ruled out. Labov (1972a) proposed therefore that speakers were responding to the fact that certain variants had taken on a stigmatized indexicality within the greater community. In other words, members of the greater community had begun to recognize a link between a marginalized subgroup (working-class people) and certain variants that the subgroup produced. In turn, a statistical majority within that subgroup had begun to avoid those same variants in more formal contextual styles. Since each of these processes occurs with a considerable lag, Labov (1972a) proposed that the degree of avoidance of a variant in formal styles was an approximate function of how long that variant had been in circulation. Variants that show steeper frequency-reduction in formal styles are referred to as *markers*, and variants that show minimal or no frequency-reduction in formal styles are referred to as *indicators*. A third potential class of variants are referred to as *stereotypes* and constitute markers in an advanced stage of stigmatization about which overt meta-commentary circulates.

This evolutionary process applies to what Labov refers to as *changes from below*, a reference to linguistic innovations that emerge from seemingly nothing, escaping the notice of most speakers and linguists during their inception. All changes from below start as indicators, but not all indicators ever become markers, and not all markers ever become stereotypes. Furthermore, pure indicators that have zero slopes on the style continuum are actually exceedingly rare; rather, it is more helpful to conceptualize this taxonomy in relative terms: i.e., "feature A is more indicator-like or more marker-like than feature B". According to Bell (1984),

> indicators in fact rarely have absolute zero shift. Most show 10–20 percent style shift (e.g. Trudgill 1974:98), and beyond that, the slope of shift is noticeable enough to class the variable as a marker (Bell 1984:154).

Trudgill (1988), for example, presented data that implies (via my own interpretation) that *t*-glottaling in Norwich[3] was *indicator-like* for speakers born between 1875 and 1908, *marker-like* for speakers born between 1908 and 1937, and never reached stereotype status. Rather, it diffused into the speech of the middle class[4] and became socially neutral for speakers born after 1937. This is illustrated in Figure 8.1. I have duplicated Trudgill's (1988) figure and superimposed my own interpretation over it.

[3] This is not to say that the *t*-glottaling innovation is a prototypical change from below; rather, it originated from the working-class vernacular in metropolitan London. But Labovian theory is unclear about borrowed features that elusively and exclusively first enter a community through the vernacular of working-class youth. This lack of clarity extends to innovations in multiethnolect that originate from migrant languages but emerge first in the speech of children.

[4] NB that the highest social class that Trudgill sampled was the "middle middle class" (Trudgill 1972:181, 1974:60–61).

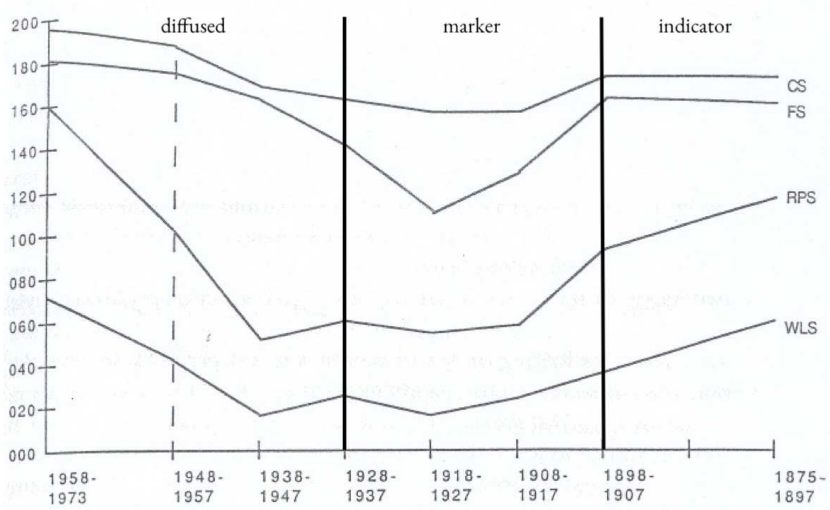


Figure 8.1: *t*-glottaling by age group and style in Norwich; taken from Trudgill (1988:45) and annotated by me. *x*-axis refers to birth cohort; *y*-axis refers to the *t*-glottaling score as calculated according to the methodology laid out in Chambers and Trudgill (1998:51–52). CS = casual style, FS = formal style, RPS = reading passage, WLS = wordlist.

Bell (1984) argues that intraspeaker variation across contextual styles comes from social stratification. The former cannot exist without the latter. For example, it is only when speakers begin to notice that other groups produce different variants that they target those variants in certain domains.

> If style variation derives from social variation, social variation comes first. So we can expect that quantitatively, some linguistic variables will have both social and style variation, some only social variation, but none style variation only (Bell 1984:151).

He elaborates that "variation on the style dimension within the speech of a single speaker derives from and echoes the variation which exists between speakers on the 'social' dimension" (1984:151). He notes that for certain variants, it may be difficult to distinguish "a casual salesman from a careful pipefitter" (Bell 1984 citing Labov 1972a:240).

This idea that style-shifting in speech production can be a *proxy* for general social salience, which itself can be a proxy for the evolution of an incipient change is relatively well-documented in the earlier First and Second Wave variationist works. For example, innovations that show mild stylistic stratification among speakers in one cohort often show steeper stylistic stratification in the next cohort. In New York, examples of this are the ingliding THOUGHT vowel and the raised ingliding TRAP vowel (Labov 2006 [1966]). Both are changes from below. THOUGHT shows the least stylistic movement among the lower working class with 52% of speakers switching to the conservative variant, and TRAP shows slightly more (75%). In an examination of the age distribution, 58 and 77% of older speakers shifted in formal speech from innovative to conservative THOUGHT and TRAP, respectively. Among younger speakers, these numbers had grown to 84 and 100%, respectively (Labov 2006 [1966]:367). The numbers imply that the formality slope is a function of evolutionary progression.

Another example of this is Cameron's (2000) investigation of direct quotation strategies in the Spanish of San Juan, Puerto Rico. The use of a recent innovation *and s/he + quotation* was favored by unskilled laborers and younger speakers like in the phrase "*y ella, '¡Ah no, mi'jo!'*" (Eng.: "*and she, 'Ah no, kiddo!'*") (2000:270, 273). Among adults, however, the production of this variant showed little stylistic variation in comparison to the two other variants, leading the author to conclude that it had indicator status. Children, however, showed more stylistic sensitivity, favoring the variant in situations of fear, humor, and surprise while disfavoring it in the remaining speech domains (2000:274).

Returning to the tie between salience and style, if a variant occurs less often in formal settings than in casual settings (and other effects are accounted for), then a reasonable explanation is that it is linked to some degree of stigmatized indexicality in the minds of a statistical majority of the speakers. It does not mean that the speakers are consciously aware of it or have full motor control over it, but it implies that the sociodemographic patterns of the variant have been absorbed, processed, and are redistributed in a systematic, albeit crude, way. In other words, it implies at a minimum subconscious awareness and partial motor control. As a case study, let us imagine a working-class Stockholmian man speaking in his 'best' formal register. Let us say he is trying to get his son into the elite Södra Latin high school and he has a meeting with the school admissions officer. His utterances would be a crude or not-so-crude production – due to individual differences in speech motor control – of the crude or not-so-crude mental representation – due to individual differences in socio-indexical and phonetic-perceptive aptitude – of the macro-community norm.

In the above scenario, the variants that would be produced at lower rates would presumably be the ones that have circulated in the community for some time. In the ensuing analysis, by means of a reading passage, I wish to simulate such a formality scenario in order to test the stylistic sensitivity of the staccato rhythm of working-class *invandrare*. I also wish to test the stylistic sensitivity of the non-staccato rhythm of working-class *svensk* speakers.

2 Research questions

As it pertains the adjustment of rhythm in more formal registers and the literature that connects stylistic sensitivity to social salience, the research question can be phrased as follows:

> *RQ: To what degree is speech rhythm stylistically sensitive in Stockholm, and what inference can be made about the maturity of its associated variants?*

3 Method

As described in Chapter 4 (and Chapter 8, pp. 128–130), speech data come from 36 male speakers, ages 24–43, from a representative sample of social-class and ethnic backgrounds. They constitute a subset of the 60 Stockholmers whom I interviewed and observed ethnographically between 2015 and 2018.

CASUAL speech was extracted from peer-group conversations when available. For speakers for whom I only had an adapted sociolinguistic interview, spontaneous speech was extracted from the interview. Criteria for treatment as 'spontaneous' was the presence of swearing, *channel cues* (Labov 1972b:113) and/or a topic that was engaging for the speaker such as danger of death or supernatural occurrences (Labov 1972b).

The two reading styles, READING and RADIO, were elicited from the same 36 speakers as described in the ensuing section.

3.1 Reading passage

Inspired by Labov's (1972b) *sociolinguistic interview*, I gave the speakers a reading passage by the name of *Cirkusen* (The Circus). They were asked read through it silently first and then read it aloud with no additional instructions.

Cirkusen was designed as a speech pathology diagnostic test (Morris & Zetterman 2011). It contains a number of improvements that the authors claim address shortcomings within the more established tests in circulation. These improvements were later ratified by a second Master's thesis that tested the passage against other more established tests that were in circulation (Wesselhoff & Gauding 2015). For this study, *Cirkusen* is also more suitable than Engstrand's (1990) *Nordanvinden och solen* (The North Wind and Sun) and the Speech Pathology diagnostic favorite *Ett svårt fall* (A Severe Case, Ingvar and Jellart 1959). For one, it is longer; *Nordanvinden och solen* and *Ett svårt fall* are much shorter with 155 and 134 syllables, respectively, providing less overall tokens. In contrast, *Cirkusen* has 367 syllables. Furthermore, *Nordanvinden och solen* uses archaic words that could key the voicing of an elderly person. Finally, *Cirkusen* has all and multiple occurrences of Central Swedish phonemes. *Nordanvinden och solen* has a comprehensive inventory but only single occurrences of all Central Swedish phonemes; *Ett svårt fall* is missing the following seven phonemes: /ʏ, yː, æ, œ, œː, ɕ, ɭ/ (Morris & Zetterman 2011:17).

Morris and Zetterman (2011) ultimately decided that the lexical inventory in *Cirkusen* was too complex, and they removed some compound words, settling on a final version that they entitled *Trapetskonstnären* (The Trapeze Artist). I decided, however, to use *Cirkusen* because I assumed that my participants would not have the same problems as readers with speech impediments. This assumption was not entirely correct. Reading aloud proved to be difficult for a number of participants who otherwise were high-functioning speakers. Certain phrases consistently rendered disfluencies, forcing me to throw out those sentences in my

Reading passage used in READING and RADIO elicitations

Cirkusen

Cirkusen var på väg! Deras plakat, med bilden av en flygande cirkusartist, var uppsatta över hela stan. Tidigare på dagen satt jag uttråkad när morfar ringde och sa:

~~– Jag har en överraskning till dig, något bra för humöret och själen. Vill du veta vad det är?~~

~~– Ja! Berätta, berätta, berätta! ropade jag förtjust.~~

~~– Du får se själv i kväll.~~

~~– Vad kul!~~

På ängen hade man rest ett enormt cirkustält. Vi var först i kön och fick platser närmast manegen. När ljuset mörknade spratt jag till av förväntan. Cirkusorkestern spelade upp och föreställningen började klockan 20.00 på pricken.

Nummer efter nummer följde: farliga tigrar, jonglörer, ormmänniskor och strutsar. Spjutkastare och sjölejon som skvätte vatten på publiken. Hästar i formation och lindansare och en magiker utklädd som en ande: den blå anden från Aladdin! Clowner som sjöng roliga låtar och en räv som hoppade genom brinnande ringar.

Höjdpunkten kom när cirkusdirektören presenterade trapetskonstnärens paradnummer. ~~Utan säkerhetslina skulle hon göra en trippelsaltomortal högst uppe under tältets tak.~~ Hon tog fart och kastade sig ut, precis som på plakatet. Jag höll hårt i morfars hand. Graciöst voltade hon genom luften men när hon greppade den andra trapetsen såg det ut som om hon skulle tappa taget.

~~– Hon faller! skrek någon.~~

I sista ögonblicket lyckades hon klamra sig fast. Stort jubel hördes och trumpeten spelade en fanfar. Dagen efter hade cirkusen dragit vidare och ängen var tom. Kvar fanns bara ett bortglömt plakat och ett minne för livet.

English translation

The Circus

The circus was on its way! Their posters, with the picture of a flying circus artist, were posted all over town. Earlier in the day I was sitting around bored when grandpa called and said:

~~– I have a surprise for you, something good for the mood and soul. Would you like to know what it is?~~

~~– Yes! Tell me, tell me, tell me, I shouted with delight.~~

~~– You will get to see yourself tonight.~~

~~– How fun!~~

An enormous circus tent had been pitched on the meadow. We were first in the queue and got seats closest to the ring. When the lights darkened I exploded with anticipation. The circus orchestra began to play, and the show began at 8 o'clock on the dot.

Number after number followed: dangerous tigers, jugglers, snake charmers, and ostriches. Javelin-throwers and sea lions that splashed water on the crowd. Horses in formation and tightrope walkers and a magician dressed like a genie: the blue genie from Aladdin! Clowns who sang funny songs and a fox that jumped through burning rings.

The climax came when the circus director presented the trapeze artist's showstopper. ~~Without a safety harness she was going to do a triple somersault at the top of the tent roof.~~ She set off and threw herself out, just like in the poster. I held on firmly to grandfather's hand. Graciously, she launched herself through the air, but when she grabbed the second trapeze, it looked like she had lost her grip.

~~– She's falling! someone screamed.~~

At the last moment she managed to regain her grip. A huge cheer was heard and the trumpets played a fanfare. The day after, the circus had moved on, and the meadow was empty. The only thing left behind was a forgotten poster and a memory for life.

Figure 8.2: My adapted version of the reading passage *Cirkusen* used for the READING and RADIO elicitations (The Circus, Morris and Zetterman 2011). The translation is provided in the bottom panel. Parts added by me to the original text are underlined. Parts excluded from the analysis are ~~struck through~~.

analysis. That said, the higher syllable inventory ($N = 364$) made it such that throwing out certain sentences did not significantly reduce token counts in the data. The passage I used and its translation are found in Figure 8.2. The parts I removed from the analysis because of regular disfluencies are struck through. Note also that the reading passage has quoted speech in it – both of a little boy and of a grandfather. These were removed from the analysis because they keyed another voice among many of the speakers and are also struck through.

I made a series of enhancements to the original *The Circus* in order to increase the inventory of certain linguistic features for the purpose of future research on certain segments. These are underlined in Figure 8.2.

3.2 Elicitation of READING and RADIO styles

Two reading elicitations were collected in order to simulate 'formal' and 'hyper-formal' speech.

Formal speech (READING)

After the casual portion of the interview, I asked the participant to read *Cirkusen* quietly to himself and then read it aloud for the recording device. No further instructions were given.

Hyper-formal speech (RADIO)

After a short break, and word lists and minimal pairs for a different study[5], I asked the participant to read the same passage again, yet this time like a radio announcer on Radio Sweden. The instructions were slightly altered depending on the speaker's racialized identification[6]:

- Speakers who self-identified as *invandrare* were prompted to read the passage as if they were a *svensk* radio presenter on *Radio Sweden*.
- Speakers who self-identified as *svensk* were prompted to read the passage as if they were a radio presenter on *Radio Sweden*.

The reason for this is as follows: With *invandrare* participants, I wanted them to specifically target the norm set by the hegemonic group (*svensk*) in order to see what variants were released from the unprompted reading style and what variants were retained. With the *svensk* participants, I had less of a specific research goal, but using *svensk* with them would be received as tautological and draw attention to the fact that ethnicity (and all of its political connotations) was a focus of the study. For the *invandrare* participants, ethnicity was already salient, because they had already discussed their family's background and home-language use prior to the reading elicitation.

[5] Word lists and minimal pairs were included in the interviews for the purpose of eventually examining segmental variation. These, however, are not analyzed in the dissertation.

[6] See pp. 86–89 for how I conceptualize race.

3.3 Segmentation and coding

As was described in Chapter 4, the recordings were transcribed by myself and a team of transcribers. The transcriptions were phonetically-timed aligned automatically via SweFA (Young 2018b; Young & McGarrah 2017). I manually corrected all segment boundaries after forced alignment. The final material for the READING style ranged between 228 and 299 vocalic elements per speaker, totaling 9 704. The final material for the RADIO style ranged between 218 and 313 vocalic elements per speaker, totaling 9 573. A summary of the internal and external properties per speaker alongside mean internal values and mean nPVIV is provided in Table 8.1. Due to the fact that word frequency, speech rate, and nPVIV occur in skewed distributions, Table 8.2 provides the same chart, albeit with median values for those three headings instead of means.

The same internal predictors are examined in this section as were examined in Chapter 6. Refer to Section 2.2 (pp. 130–141) for a detailed account of how they are operationalized. As a reminder, they are:

coding ID: ACCENT_1_LONG
coding ID: ACCENT_1_SHORT
coding ID: ACCENT_2_LONG
coding ID: ACCENT_2_SHORT
coding ID: UNSTRESSED_LONG
coding ID: UNSTRESSED_SHORT
coding ID: PHRASE_FINAL
coding ID: CODA_R
coding ID: SPEECH_RATE
coding ID: LEXICAL_FREQ

Three different analyses are run on different external predictors, Therefore, the coding of these predictors will be reviewed at the start of each Analysis subsection.

4 Analysis

This analysis consists of three components. First, I provide visualizations of the relationship between social predictors and nPVIV for each contextual style. I then build mixed-effects linear regression models to test the visual impression gained from each. AGE, RACIALIZATION, and the first principle component of my six social-class criteria (PC1_SOCIAL_CLASS) are the social predictors tested here. Second, I validate these results against a random-forest variable importance analysis for the individual social and social-network predictors. In the third and final analysis, I build a mixed-effects linear regression model with STYLE as one of the predictors in order to validate the findings in the first analysis and isolate the effects of formality from the mechanical effects of reading aloud.

Group	Pseudonym	Profession	*n* accent-1 long vowels	*n* accent-1 short vowels	*n* accent-2 long vowels	*n* accent-2 short vowels	*n* unstressed long vowels	*n* unstressed short vowels	*n* phrase-final vowels	*n* /r/ in coda position	mean speech rate in milliseconds	mean lexical frequency
			Internal factors (CASUAL) - Summary									
‘invandrare’, social class: working	Reman	social worker	55	37	12	29	159	254	26	29	150	7 641
	Malik	cook at cafeteria, street food	77	55	18	34	117	253	48	7	147	7 019
	Max	youth worker	149	86	44	71	318	565	57	62	134	7 004
	Antonio	caretaker, ex drug dealer	166	96	73	124	307	659	63	65	160	5 927
	Hayder	caretaker, bouncer	85	59	38	77	224	403	36	24	160	6 243
	Dawit	group home assistant	195	82	73	93	361	761	78	56	141	5 583
	Solomon	youth worker	150	120	60	112	370	820	61	74	137	7 776
	Murad	group manager, furniture retail	98	64	36	58	180	389	43	14	137	5 190
	Abel	photographer, personal trainer	110	53	30	61	201	383	52	34	149	7 800
	Mateo	waiter, bartender	93	63	33	48	278	432	60	43	145	7 502
	Reza	waiter, napropathy student	120	55	46	39	230	457	39	28	154	5 634
	Mezdar	youth worker (certified)	118	45	21	51	185	369	47	46	150	6 893
	Jocke	athletic coach, personal trainer	194	100	62	100	450	826	90	70	144	7 604
	Sohrab	web architect	149	67	37	61	298	548	67	48	151	6 529
	Shorty	social worker	89	82	22	45	195	325	62	30	137	8 545
	Kevin	organisational consultant	137	67	40	64	316	597	54	42	140	7 532
upper	Tarik	field assistant; athletic coach	108	89	37	56	200	454	51	39	151	5 168
	Majeed	doctor	133	85	32	55	278	544	59	36	134	7 952
	Parviz	business owner, physician	116	40	33	54	253	349	45	34	148	8 989
‘svensk’, social class: working	Per	demolition operative	66	28	27	36	205	363	43	27	153	9 555
	Rasmus	chef at major hotel	74	48	39	51	213	361	57	25	141	7 702
	Johnny	waiter, bartender	115	88	51	58	251	443	63	47	166	6 541
	Sven	team leader, security firm	50	38	29	36	135	324	29	38	151	8 703
	Paul	self-employed, renovations	91	67	32	46	240	414	62	54	156	7 528
	Jesper	waiter	61	49	35	64	225	427	45	38	132	7 672
	Richie	pest exterminator	112	77	50	44	284	486	37	57	149	9 353
	Thomas	middle office associate, hedgefund	173	68	35	65	376	648	54	62	140	8 233
	Nils	supermarket clerk	100	62	23	51	208	456	44	34	152	8 233
	August	section manager, IT	66	58	29	33	251	352	30	39	149	8 025
	Jonte	middle school teacher, special ed	113	41	55	60	263	513	47	34	144	6 249
	Martin	massage therapist, former teacher	101	37	38	31	287	422	41	23	148	8 346
	Gunnar	high school teacher	118	51	30	43	235	390	52	38	144	7 395
upper	Joseph	communications strategist/PR	84	47	24	37	148	296	38	23	178	7 675
	Jan-Bertel	portfolio manager, hedgefund	126	65	39	44	292	433	69	28	132	6 478
	Johan	chief marketing officer	38	17	29	34	82	173	20	22	162	9 098
	Jan-Axel	trader, hedgefund	92	48	45	45	154	347	61	35	170	7 633
		Means:	*109*	*62*	*38*	*56*	*244*	*451*	*51*	*39*	*148*	*7 415*
		Percentages:	19%	11%	6%	10%	42%	77%	9%	7%		

Pseudonym	*n* accent-1 long vowels	*n* accent-1 short vowels	*n* accent-2 long vowels	*n* accent-2 short vowels	*n* unstressed long vowels	*n* unstressed short vowels	*n* phrase-final vowels	*n* /r/ in coda position	mean speech rate in milliseconds	mean lexical frequency
	Internal factors (READING) - Summary									
Reman	39	21	23	29	67	200	20	33	182	2 350
Malik	51	44	32	47	52	220	40	34	160	2 205
Max	46	23	33	45	74	262	30	41	161	2 076
Antonio	46	37	33	43	65	223	26	35	178	2 592
Hayder	55	32	42	42	47	236	25	35	172	2 424
Dawit	43	23	37	44	81	238	35	36	151	2 368
Solomon	35	18	28	39	84	228	27	26	151	2 358
Murad	49	29	35	40	74	252	27	36	154	2 350
Abel	39	31	33	46	92	237	30	34	160	2 084
Mateo	41	29	31	37	97	240	20	30	145	2 647
Reza	40	24	27	35	71	212	20	36	168	2 205
Mezdar	50	28	28	31	58	217	38	30	178	1 931
Jocke	54	29	30	34	67	241	25	31	154	2 519
Sohrab	45	12	22	40	81	214	24	24	161	2 235
Shorty	51	31	32	51	66	242	28	30	160	1 908
Kevin	39	35	33	33	70	225	22	33	176	2 445
Tarik	47	18	28	28	71	216	21	34	174	2 332
Majeed	47	20	23	38	71	222	32	36	155	2 221
Parviz	43	25	31	34	57	200	29	39	175	2 481
Per	47	29	34	32	81	217	24	29	171	2 621
Rasmus	38	27	32	33	74	230	21	36	167	1 913
Johnny	48	26	39	47	84	227	30	30	174	2 548
Sven	37	27	39	39	87	247	25	32	160	2 616
Paul	39	22	35	39	73	217	26	32	161	2 270
Jesper	35	29	28	43	103	234	30	38	155	2 449
Richie	32	24	33	39	82	222	19	34	164	2 373
Thomas	45	26	32	27	67	257	20	36	160	2 501
Nils	37	22	28	35	94	227	22	28	168	2 236
August	48	27	43	35	83	241	22	29	176	2 185
Jonte	29	20	32	33	96	223	18	32	157	2 160
Martin	26	12	22	23	84	206	16	32	164	2 457
Gunnar	36	22	29	38	85	255	25	34	149	2 661
Joseph	39	20	23	29	80	222	27	36	168	2 488
Jan-Bertel	42	31	34	35	78	236	34	35	148	2 111
Johan	35	21	25	43	91	236	20	32	159	2 274
Jan-Axel	48	27	32	30	55	219	28	38	180	2 411
Means:	*42*	*26*	*31*	*37*	*76*	*229*	*26*	*33*	*164*	*2 333*
Percentages:	16%	9%	12%	14%	28%	85%	10%	12%		

Pseudonym	*n* accent-1 long vowels	*n* accent-1 short vowels	*n* accent-2 long vowels	*n* accent-2 short vowels	*n* unstressed long vowels	*n* unstressed short vowels	*n* phrase-final vowels	*n* /r/ in coda position	mean speech rate in milliseconds	mean lexical frequency
	Internal factors (RADIO) - Summary									
Reman	37	11	29	27	69	200	24	32	179	2 442
Malik	51	38	33	53	72	220	34	34	165	1 857
Max	57	30	42	44	73	262	27	42	152	2 423
Antonio	45	32	36	43	70	223	35	34	180	2 111
Hayder	39	17	23	38	78	236	27	30	171	1 954
Dawit	37	20	37	40	84	238	38	27	155	1 940
Solomon	37	36	21	50	81	228	31	36	160	1 996
Murad	38	24	32	36	88	252	26	31	148	2 251
Abel	40	22	35	42	96	237	39	30	155	2 443
Mateo	35	22	33	27	93	240	25	28	146	2 218
Reza	30	24	33	35	77	212	20	30	172	2 743
Mezdar	52	25	29	43	63	217	36	36	164	2 212
Jocke	44	20	33	31	71	241	25	33	158	2 512
Sohrab	42	16	19	36	90	214	22	29	168	2 265
Shorty	54	29	28	47	89	242	26	33	163	2 372
Kevin	35	31	30	36	84	225	20	31	178	2 305
Tarik	38	14	26	25	84	216	17	37	179	2 792
Majeed	41	21	27	43	83	222	29	39	155	2 162
Parviz	37	21	32	25	59	200	29	38	180	2 088
Per	42	26	34	45	93	217	33	36	166	2 175
Rasmus	40	22	41	39	69	230	21	36	172	2 128
Johnny	43	25	41	48	78	227	31	25	173	2 563
Sven	35	29	45	42	86	247	34	33	158	2 422
Paul	31	20	35	33	79	217	22	34	166	2 359
Jesper	34	25	35	34	92	234	33	33	163	2 636
Richie	29	23	30	34	81	222	19	32	164	2 580
Thomas	37	23	34	27	70	257	17	37	173	2 266
Nils	35	14	27	33	79	227	30	32	185	1 811
August	35	26	44	27	90	241	21	35	179	2 642
Jonte	35	18	29	38	103	223	21	29	151	2 344
Martin	13	10	19	28	102	206	16	28	166	2 590
Gunnar	38	18	32	32	76	255	29	29	146	2 316
Joseph	37	17	34	32	82	222	29	41	159	2 322
Jan-Bertel	47	28	37	35	72	236	30	36	151	1 929
Johan	41	20	26	28	87	236	21	37	157	2 223
Jan-Axel	39	21	34	38	64	219	22	28	173	2 455
Means:	*39*	*23*	*32*	*37*	*81*	*229*	*27*	*33*	*165*	*2 301*
Percentages:	15%	9%	12%	14%	30%	86%	10%	12%		

Pseudonym	*n* vocalic intervals (CASUAL)	mean nPVIV (CASUAL)	*n* vocalic intervals (READING)	mean nPVIV (READING)	*n* vocalic intervals (RADIO)	mean nPVIV (RADIO)	Group
	Speech rhythm (ALL STYLES)						
Reman	340	55.8	228	65.2	218	63.5	*‘invandrare’*, social class: working
Malik	333	48.0	264	49.4	275	52.3	
Max	742	48.8	299	49.4	313	48.9	
Antonio	849	45.6	270	49.7	262	51.1	
Hayder	539	54.8	278	54.1	267	68.3	
Dawit	961	51.3	282	55.9	267	54.3	
Solomon	993	59.0	272	63.6	265	58.4	
Murad	504	45.2	294	51.8	282	52.7	
Abel	500	54.6	293	56.5	287	55.5	
Mateo	585	60.6	285	62.7	277	65.6	
Reza	583	51.9	253	52.2	245	60.7	
Mezdar	473	52.4	256	51.5	271	65.7	
Jocke	1 065	57.0	274	59.8	263	63.3	
Sohrab	714	63.5	256	67.1	257	65.5	
Shorty	455	52.1	287	49.4	302	52.3	
Kevin	761	58.1	259	63.1	258	61.5	
Tarik	573	49.2	255	63.0	245	60.9	upper
Majeed	693	65.4	258	69.0	272	65.6	
Parviz	507	57.5	241	60.2	229	59.0	
Per	453	78.3	267	57.4	281	63.0	*‘svensk’*, social class: working
Rasmus	478	72.3	268	64.2	268	75.5	
Johnny	591	64.6	278	60.3	259	59.8	
Sven	379	61.5	290	61.1	283	55.8	
Paul	543	74.3	258	57.4	245	61.9	
Jesper	541	64.8	285	61.2	279	59.4	
Richie	637	60.4	258	65.6	248	71.1	
Thomas	822	65.6	286	67.1	256	63.7	
Nils	565	57.2	267	71.1	259	72.8	
August	482	67.0	284	63.7	275	63.6	
Jonte	640	71.2	264	65.1	273	67.5	
Martin	557	91.0	245	87.7	239	87.2	
Gunnar	517	60.0	286	64.3	265	69.5	
Joseph	378	56.0	256	63.9	278	65.5	upper
Jan-Bertel	584	61.6	277	67.1	275	65.7	
Johan	220	57.5	272	74.5	282	68.1	
Jan-Axel	443	51.7	259	55.6	253	61.9	
Means:	*21 000*	*59.6*	*9 704*	*61.1*	*9 573*	*62.7*	

Table 8.1: Overview of full Stockholm dataset for all contextual speech styles using <u>means</u>: CASUAL, READING, and RADIO. Internal predictors: total counts for accent 1 and 2 long and short vowels, unstressed long and short vowels, vowels in phrase-final position, rhotics in coda position, mean speech rate in milliseconds, mean lexical frequency. The final six columns are the *n* vocalic intervals *n* alongside the mean nPVIV for CASUAL, READING, and RADIO. Speakers are grouped by the racialized *invandrare/svensk* binary and ranked according to Social class (PC1).

Group	Social class	Pseudonym	Profession	*n* accent-1 long vowels	*n* accent-1 short vowels	*n* accent-2 long vowels	*n* accent-2 short vowels	*n* unstressed long vowels	*n* unstressed short vowels	*n* phrase-final vowels	*n* /r/ in coda position	median speech rate in milliseconds	median lexical frequency
				Internal factors (CASUAL) - Summary									
'invandrare'	working	Reman	social worker	55	37	12	29	159	254	26	29	142	3 145
		Malik	cook at cafeteria, street food	77	55	18	34	117	253	48	7	144	2 858
		Max	youth worker	149	86	44	71	318	565	57	62	125	2 867
		Antonio	caretaker, ex drug dealer	166	96	73	124	307	659	63	65	157	1 105
		Hayder	caretaker, bouncer	85	59	38	77	224	403	36	24	154	1 740
		Dawit	group home assistant	195	82	73	93	361	761	78	56	135	1 608
		Solomon	youth worker	150	120	60	112	370	820	61	74	132	1 818
		Murad	group manager, furniture retail	98	64	36	58	180	389	43	14	131	1 110
		Abel	photographer, personal trainer	110	53	30	61	201	383	52	34	143	3 496
		Mateo	waiter, bartender	93	63	33	48	278	432	60	43	135	2 283
		Reza	waiter, napropathy student	120	55	46	39	230	457	39	28	148	2 688
		Mezdar	youth worker (certified)	118	45	21	51	185	369	47	46	142	1 749
		Jocke	athletic coach, personal trainer	194	100	62	100	450	826	90	70	135	2 741
		Sohrab	web architect	149	67	37	61	298	548	67	48	144	2 240
		Shorty	social worker	89	82	22	45	195	325	62	30	125	3 527
		Kevin	organisational consultant	137	67	40	64	316	597	54	42	133	2 642
		Tarik	field assistant; athletic coach	108	89	37	56	200	454	51	39	143	1 463
		Majeed	doctor	133	85	32	55	278	544	59	36	125	3 381
	upper	Parviz	business owner, physician	116	40	33	54	253	349	45	34	139	3 344
'svensk'	working	Per	demolition operative	66	28	27	36	205	363	43	27	147	4 470
		Rasmus	chef at major hotel	74	48	39	51	213	361	57	25	134	2 900
		Johnny	waiter, bartender	115	88	51	58	251	443	63	47	163	2 690
		Sven	team leader, security firm	50	38	29	36	135	324	29	38	141	3 494
		Paul	self-employed, renovations	91	67	32	46	240	414	62	54	147	2 265
		Jesper	waiter	61	49	35	64	225	427	45	38	126	2 172
		Richie	pest exterminator	112	77	50	44	284	486	37	57	143	5 324
		Thomas	middle office associate, hedgefund	173	68	35	65	376	648	54	62	134	3 145
		Nils	supermarket clerk	100	62	23	51	208	456	44	34	145	3 331
		August	section manager, IT	66	58	29	33	251	352	30	39	141	3 673
		Jonte	middle school teacher, special ed	113	41	55	60	263	513	47	34	140	1 268
		Martin	massage therapist, former teacher	101	37	38	31	287	422	41	23	140	3 468
		Gunnar	high school teacher	118	51	30	43	235	390	52	38	137	2 018
		Joseph	communications strategist/PR	84	47	24	37	148	296	38	23	170	2 705
		Jan-Bertel	portfolio manager, hedgefund	126	65	39	44	292	433	69	28	121	1 583
		Johan	chief marketing officer	38	17	29	34	82	173	20	22	151	2 824
	upper	Jan-Axel	trader, hedgefund	92	48	45	45	154	347	61	35	154	2 642
			Medians:	*109*	*61*	*36*	*51*	*238*	*425*	*52*	*37*	*139*	*2 601*
			Percentages:	19%	11%	6%	10%	42%	77%	9%	7%		

Pseudonym	*n* accent-1 long vowels	*n* accent-1 short vowels	*n* accent-2 long vowels	*n* accent-2 short vowels	*n* unstressed long vowels	*n* unstressed short vowels	*n* phrase-final vowels	*n* /r/ in coda position	median speech rate in milliseconds	median lexical frequency
	Internal factors (READING) - Summary									
Reman	39	21	23	29	67	200	20	33	177	23
Malik	51	44	32	47	52	220	40	34	156	61
Max	46	23	33	45	74	262	30	41	156	65
Antonio	46	37	33	43	65	223	26	35	176	136
Hayder	55	32	42	42	47	236	25	35	170	70
Dawit	43	23	37	44	81	238	35	36	148	41
Solomon	35	18	28	39	84	228	27	26	145	80
Murad	49	29	35	40	74	252	27	36	150	66
Abel	39	31	33	46	92	237	30	34	155	34
Mateo	41	29	31	37	97	240	20	30	141	122
Reza	40	24	27	35	71	212	20	36	161	72
Mezdar	50	28	28	31	58	217	38	30	174	34
Jocke	54	29	30	34	67	241	25	31	150	90
Sohrab	45	12	22	40	81	214	24	24	159	51
Shorty	51	31	32	51	66	242	28	30	157	30
Kevin	39	35	33	33	70	225	22	33	174	58
Tarik	47	18	28	28	71	216	21	34	173	35
Majeed	47	20	23	38	71	222	32	36	153	70
Parviz	43	25	31	34	57	200	29	39	172	118
Per	47	29	34	32	81	217	24	29	173	87
Rasmus	38	27	32	33	74	230	21	36	166	64
Johnny	48	26	39	47	84	227	30	30	172	66
Sven	37	27	39	39	87	247	25	32	157	80
Paul	39	22	35	39	73	217	26	32	162	49
Jesper	35	29	28	43	103	234	30	38	153	68
Richie	32	24	33	39	82	222	19	34	162	72
Thomas	45	26	32	27	67	257	20	36	158	123
Nils	37	22	28	35	94	227	22	28	166	35
August	48	27	43	35	83	241	22	29	176	80
Jonte	29	20	32	33	96	223	18	32	153	34
Martin	26	12	22	23	84	206	16	32	163	68
Gunnar	36	22	29	38	85	255	25	34	150	70
Joseph	39	20	23	29	80	222	27	36	160	80
Jan-Bertel	42	31	34	35	78	236	34	35	142	52
Johan	35	21	25	43	91	236	20	32	156	32
Jan-Axel	48	27	32	30	55	219	28	38	175	131
Medians:	*43*	*26*	*32*	*38*	*76*	*227*	*25*	*34*	*160*	*62*
Percentages:	16%	9%	12%	14%	28%	85%	10%	12%		

Pseudonym	*n* accent-1 long vowels	*n* accent-1 short vowels	*n* accent-2 long vowels	*n* accent-2 short vowels	*n* unstressed long vowels	*n* unstressed short vowels	*n* phrase-final vowels	*n* /r/ in coda position	median speech rate in milliseconds	median lexical frequency
	Internal factors (RADIO) - Summary									
Reman	37	11	29	27	69	200	24	32	179	28
Malik	51	38	33	53	72	220	34	34	163	72
Max	57	30	42	44	73	262	27	42	150	92
Antonio	45	32	36	43	70	223	35	34	178	59
Hayder	39	17	23	38	78	236	27	30	169	60
Dawit	37	20	37	40	84	238	38	27	152	34
Solomon	37	36	21	50	81	228	31	36	157	87
Murad	38	24	32	36	88	252	26	31	146	59
Abel	40	22	35	42	96	237	39	30	149	48
Mateo	35	22	33	27	93	240	25	28	143	42
Reza	30	24	33	35	77	212	20	30	167	89
Mezdar	52	25	29	43	63	217	36	36	162	29
Jocke	44	20	33	31	71	241	25	33	158	68
Sohrab	42	16	19	36	90	214	22	29	169	58
Shorty	54	29	28	47	89	242	26	33	161	39
Kevin	35	31	30	36	84	225	20	31	175	62
Tarik	38	14	26	25	84	216	17	37	175	92
Majeed	41	21	27	43	83	222	29	39	148	70
Parviz	37	21	32	25	59	200	29	38	178	34
Per	42	26	34	45	93	217	33	36	168	42
Rasmus	40	22	41	39	69	230	21	36	169	64
Johnny	43	25	41	48	78	227	31	25	167	58
Sven	35	29	45	42	86	247	34	33	148	70
Paul	31	20	35	33	79	217	22	34	167	62
Jesper	34	25	35	34	92	234	33	33	157	117
Richie	29	23	30	34	81	222	19	32	162	141
Thomas	37	23	34	27	70	257	17	37	165	56
Nils	35	14	27	33	79	227	30	32	183	23
August	35	26	44	27	90	241	21	35	181	51
Jonte	35	18	29	38	103	223	21	29	148	60
Martin	13	10	19	28	102	206	16	28	159	65
Gunnar	38	18	32	32	76	255	29	29	147	23
Joseph	37	17	34	32	82	222	29	41	155	28
Jan-Bertel	47	28	37	35	72	236	30	36	148	51
Johan	41	20	26	28	87	236	21	37	150	32
Jan-Axel	39	21	34	38	64	219	22	28	167	62
Medians:	*38*	*22*	*33*	*36*	*81*	*227*	*27*	*33*	*161*	*58*
Percentages:	15%	9%	12%	14%	30%	86%	10%	12%		

Pseudonym	*n* vocalic intervals (CASUAL)	median nPVIV (CASUAL)	*n* vocalic intervals (READING)	median nPVIV (READING)	*n* vocalic intervals (RADIO)	median nPVIV (RADIO)	Social class	Group
	Speech rhythm (ALL STYLES)							
Reman	340	48.5	228	53.7	218	60.6	working	'invandrare'
Malik	333	42.0	264	45.0	275	50.8		
Max	742	40.6	299	41.6	313	47.4		
Antonio	849	38.4	270	45.2	262	46.2		
Hayder	539	43.7	278	44.6	267	52.6		
Dawit	961	46.6	282	52.3	267	48.3		
Solomon	993	49.7	272	52.7	265	51.8		
Murad	504	38.9	294	46.3	282	46.8		
Abel	500	48.9	293	52.0	287	48.8		
Mateo	585	49.9	285	56.1	277	58.2		
Reza	583	48.5	253	46.6	245	57.1		
Mezdar	473	46.2	256	47.7	271	58.6		
Jocke	1 065	46.3	274	53.6	263	56.7		
Sohrab	714	53.9	256	53.6	257	60.6		
Shorty	455	45.8	287	42.3	302	50.2		
Kevin	761	52.3	259	57.9	258	58.8		
Tarik	573	44.2	255	60.3	245	53.0		
Majeed	693	53.3	258	63.6	272	60.0		
Parviz	507	47.8	241	57.1	229	50.8	upper	
Per	453	70.1	267	55.4	281	56.3	working	'svensk'
Rasmus	478	55.0	268	60.1	268	69.7		
Johnny	591	53.2	278	59.6	259	52.5		
Sven	379	50.7	290	56.8	283	50.2		
Paul	543	64.4	258	55.2	245	58.4		
Jesper	541	55.3	285	55.7	279	54.2		
Richie	637	51.5	258	63.3	248	62.6		
Thomas	822	52.0	286	62.6	256	57.3		
Nils	565	47.0	267	57.1	259	61.5		
August	482	54.5	284	65.4	275	58.6		
Jonte	640	59.0	264	63.7	273	63.7		
Martin	557	73.4	245	87.5	239	75.9		
Gunnar	517	53.5	286	57.0	265	61.9		
Joseph	378	50.8	256	56.0	278	58.8		
Jan-Bertel	584	55.6	277	63.1	275	58.6		
Johan	220	53.4	272	60.9	282	62.0		
Jan-Axel	443	44.6	259	47.6	253	51.6	upper	
Medians:	*21 000*	*49.8*	*9 704*	*54.5*	*9 573*	*55.7*		

Table 8.2: A reproduction of Table 8.1, albeit with median values instead of means.

	Model 1		Model 2		Model 3	
Response variable	nPVIV† (casual)		nPVIV† (reading)		nPVIV† (radio)	
(Intercept)	2.3 (9.5)		3.2 (11.8)		−18.4 (12.2)	
Internal predictors		vif		vif		vif
accent_1_long·*yes*	23.7 (1.0)***	1.4	23.9 (2.0)***	1.5	22.8 (2.1)***	1.4
accent_1_short·*yes*	14.6 (1.2)***	1.2	13.8 (1.9)***	1.1	14.1 (2.0)***	1.1
accent_2_long·*yes*	18.3 (1.5)***	1.2	22.7 (2.2)***	1.5	23.5 (2.3)***	1.4
accent_2_short·*yes*	10.7 (1.2)***	1.2	10.6 (1.6)***	1.2	8.1 (1.7)***	1.2
unstressed_long·*yes*	8.6 (1.0)***	1.8	7.8 (1.8)***	2.1	7.3 (1.9)***	2.0
unstressed_short·*yes*	15.5 (1.1)***	1.7	18.1 (1.8)***	1.5	19.9 (1.9)***	1.4
coda_rhotic·*yes*	7.9 (1.3)***	1.0	10.5 (2.5)***	1.1	9.7 (2.8)***	1.0
phrase_finality·*yes*	6.9 (1.1)***	1.0	6.0 (1.7)***	1.0	7.3 (1.8)***	1.0
log(lexical_frequency)	−0.3 (0.1)**	1.1	−0.1 (0.2)	1.1	0.2 (0.2)	1.1
log(speech_rate)	5.9 (0.8)***	1.1	2.4 (1.7)	1.1	6.7 (1.8)***	1.1
Social predictors		vif		vif		vif
age	0.7 (0.3)**	1.5	0.5 (0.2)*	1.5	0.5 (0.2)*	1.5
racialization·*invandrare*	−23.0 (5.3)***	4.8	−5.2 (4.8)	4.8	−5.0 (4.8)	4.8
class	−0.2 (0.1)**	3.1	0.1 (0.1)	3.1	0.0 (0.1)	3.1
racialization·*invandrare*:class	0.3 (0.1)**	4.1	0.0 (0.1)	4.1	0.0 (0.1)	4.1
aic	218 844.6		98 480.2		97 640.8	
bic	218 987.7		98 609.4		97 769.8	
Log Likelihood	−109 404.3		−49 222.1		−48 802.4	
Num. obs.	21 000		9 704		9 573	
Num. groups: vowel	3 099		632		569	
Num. groups: speaker	36		36		36	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$ †The nPVIV percentages are multiplied by 100 in this dataset

Table 8.3: Mixed-effects linear regression models for each of the contextual styles: casual (Model 1), reading (Model 2), and radio (Model 3). nPVIV is the response variable. All models contain the same internal predictors and the same social predictors: age, racialization, and social class (defined as the first principle component PC1 of SEI, income, education, parental SEI, parental education, and *Mosaic* taste). For categorical predictors, the reference category is in italics. Coefficients are indicated in the center column, standard errors in the parentheses, and variance inflation factors (VIF) to the right.

4.1 Comparing regression models between contextual styles

Only a subset of social predictors are examined in this first analysis, namely age, race, and the principle component 1 (PC1) of the six social-class features (income, SEI, education, parental SEI, parental education, *Mosaic* taste). Refer to pages 144–153 for a detailed account of how they are operationalized. They are:

coding id: age
coding id: racialization
coding id: pc1_social_class

Figure 8.3 contains a visualization for the data by plotting nPVIV in the three styles against racialization and social class. They show a reduction in nPVIV stratification as speech formality increases. (Figure 8.4 below will be discussed further along in this chapter. It contains a visualization for the data by plotting nPVIV in the three styles against racialization and age).

In order to test the visual trends in Figure 8.3, I ran three mixed-effects linear regression models – one for casual, reading, and radio each – in *lme4* in *R* with the following call framework:

nPVIV ~ ACCENT_1_LONG + ACCENT_1_SHORT + ACCENT_2_LONG + ACCENT_2_SHORT + UNSTRESSED_LONG + UNSTRESSED_SHORT + CODA_R + PHRASE_FINAL + LOG(LEXICAL_FREQ) + LOG(SPEECH_RATE) + AGE + RACIALIZATION*PCI_SOCIAL_CLASS + (1|SPEAKER) + (1|VOWEL)

Table 8.3 shows the three model outputs. Table 8.3 Model 1 for CASUAL is identical to Model 2 from Tables 6.12 (p. 152) and 6.13 (p. 157). Models 2 and 3 show a decrease in strength and a lost in significance for racialization, social class, and their interaction. The coefficient for racialization decreases in absolute value from 23.0 to 5.2 to 5.0. The coefficient for social class decreases in absolute value from 0.2 to 0.1 to 0.0. The coefficient for the interaction of racialization and class decreases from 0.3 to 0.0 to 0.0. READING and RADIO show a slight decrease in strength for age; however, the statistical significance for age is maintained for every speech style.

INTERPRETING THE EFFECTS OF RACIALIZATION AND CLASS The trends in the regression outputs on Table 8.3 reflect the visualizations in Figure 8.3. Speakers appear to be targeting a shared community norm. A clear interaction between racialization and class in CASUAL results in opposing effects for nPVIV, something discussed at length in Chapter 6. This effect disappears in READING and is replaced by a uniform slope for both racialized groups that resembles the slope of the *invandrare* speakers in CASUAL. The difference, however, is that the *y*-intercept for the *invandrare* speakers in READING is approximately four points higher than their slope in CASUAL. The *svensk* slope in READING is nearly identical to the *invandrare* slope in READING, albeit with an even higher *y*-intercept. Interestingly, it shows that the group that style-shifted the *most* is the *svensk* working class. Moving to RADIO, its slopes resemble those of READING, yet even flatter, confirming that formality has a neutralizing effect on nPVIV variation.

From the regression outputs and Figure 8.3, three main trends stand out. First, the *invandrare* speakers are predicted to shift their speech rhythm toward less staccato when moving from CASUAL to READING, and the lower classes continue this shift from READING to RADIO whereas the upper-class *invandrare* speakers remain stable between READING and RADIO and are responsible for the flattening of that slope.

Second, *svensk* working-class speakers seem to be 'doing the most work' (seen in the visualization; this is not clear from the regression output). They shift by more than seven nPVIV points between CASUAL and READING.

Third, the 'intermediate' nPVIV range of upper-class CASUAL speech appears to be the target for the lower social groups in READING and RADIO. All three trends imply that rhythm carries some degree of social meaning in Stockholm.

INTERPRETING AGE Figure 8.4 shows that regardless of style, age has a relatively stable effect. Younger corresponds with lower nPVIV, and older corresponds with higher nPVIV.

Figure 8.3: Scatterplots of nPVIV in CASUAL, READING, and RADIO according to racialization and social class (defined as the principle component (PC1) of SEI, income, education, parental SEI, parental education, and *Mosaic* taste). *Svensk* speakers are in blue; *invandrare* speakers in red. Least-squares regression lines are plotted with confidence-interval halos.

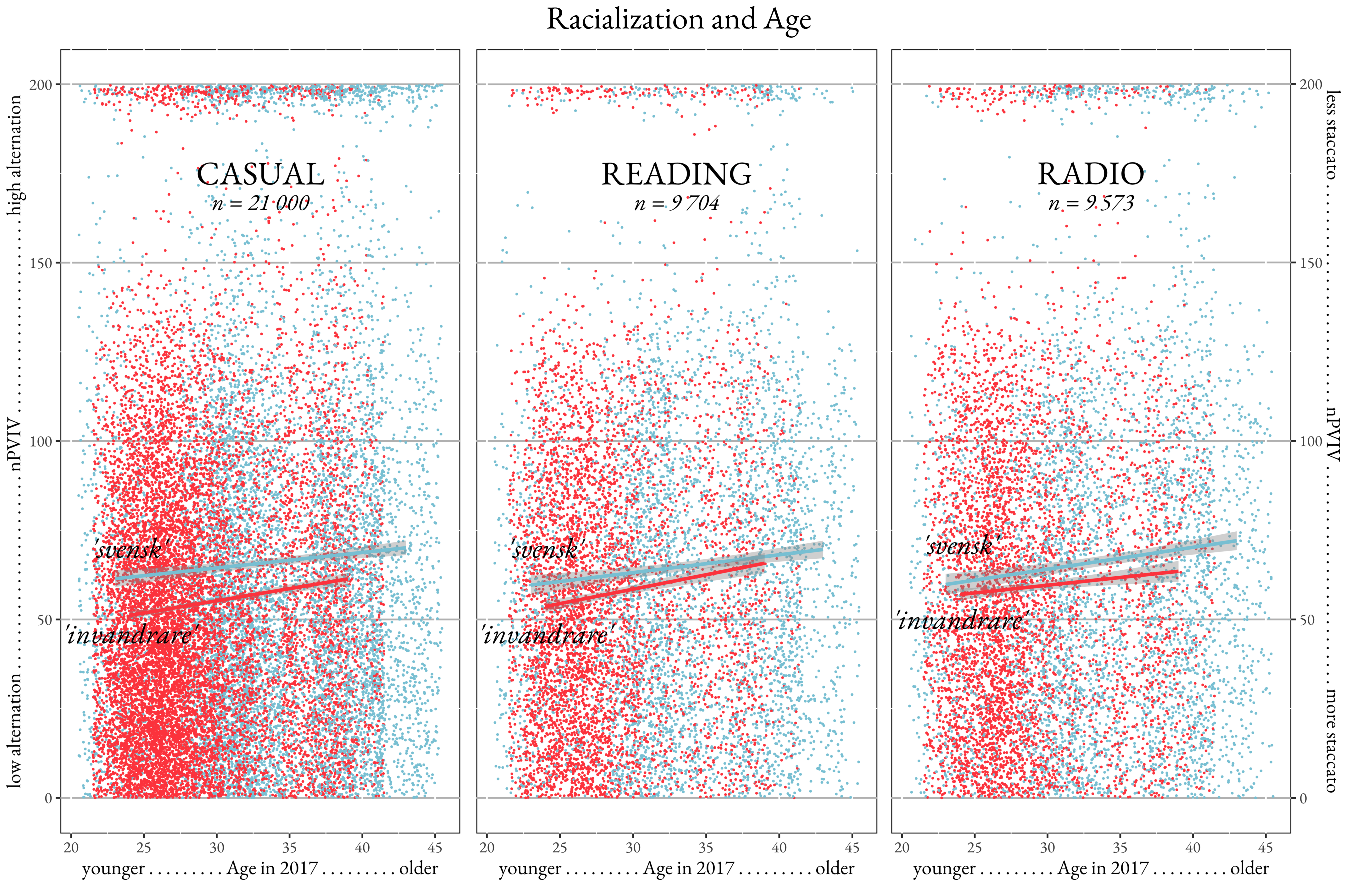


Figure 8.4: Scatterplots of nPVIV in CASUAL, READING, and RADIO according to age in 2017. *Svensk* speakers are in blue; *invandrare* speakers in red. Least-squares regression lines are plotted with confidence-interval halos.

One small discrepancy between *svensk* and *invandrare* speakers is that *svensk* speakers have the most stable age trends. All three styles showing nearly identical least-regression smoothers. On the other hand, *invandrare* speakers show less stability in slope. Their y-intercept climbs steadily upward from CASUAL to READING, implying that those speakers, especially the young ones, might generally associate higher alternation with formality.

From the regression outputs in Table 8.3, two main trends stand out. First, age has a stable, albeit mild, main effect on nPVIV, regardless of speech style, with younger speakers correlating with lower nPVIV and older speakers correlating with higher nPVIV. This implies that there may be a strand of staccato rhythm that is diffusing in from multiethnolect. This possibility is examined more closely in the next chapter by using social-network data.

Second, among *invandrare* speakers, there appears to be some association with higher nPVIV and formality, particularly when one examines the difference between CASUAL and RADIO for the 25-year-old range. Aside from that, no clear differences in style-shifting between the age groups are visible in these diagrams, which implies that age alone has no main effect. In the next chapter, however, age and its interaction with race and class are explored more in connection to stylistic sensitivity.

4.2 Validating the analysis against all social predictors

A major challenge presented by this research question is that 36 speakers are tested to make linguistic generalizations about an urban population of 1 562 136. To compensate for such a small sample size, I attempted to build unusually deep and complex social profiles for each speaker. In Chapter 6, I then attempted to empirically catalogue speakers by means of a social class metric that found the most common component of six key class factors.

Nonetheless, one always wonders with these sorts of analyses if class artifacts haven't been cherry-picked such that they render the 'right' results – especially when the significance levels are as high as in Table 8.3 (p. 221). The aim with this next analysis is to add another layer of transparency to the dissertation. In this section, I test every class artifact and every social-network quality to see whether their trends resemble the results in the prior analysis.

Introducing random forests

Random forest and *bagging-ensemble algorithms* are relatively new Decision Sciences tools that were invented by Breiman (2001). Tagliamonte and Baayen (2012) first introduced the technique to the field of Linguistics, and since then, a number of prominent studies have followed suit (Baayen, Endresen, Janda, Makarova, & Nesset 2013; Brown, Winter, Idemaru, & Grawunder 2014; Kirkham & Moore 2016; Sadat, Martin, Costa, & Alario 2014).

The toolkit makes use of advanced computer processing power to model iteration after iteration of a dataset while making small changes to each iteration, such as the removal of a predictor, in order to see how the response variable changes. In that sense, it is a nonparametric analysis technique that makes no assumptions about the distribution of the population from which a sample was drawn (Tagliamonte & Baayen 2012:136). So unlike in a linear regression analysis, the predictors do not have to be orthogonal, and the normality criterion does not have to be met. This is particularly useful for sociolinguistic analyses such as the one at hand. The Stockholm dataset has a relatively high number of social predictors ($n = 15$) and a relatively low number of participants ($n = 36$ from an urban population of $N = 1\,562\,136$). This is popularly known as the *small n/ large p problem* (Strobl, Boulesteix, Kneib, Augustin, & Zeileis 2008; Tagliamonte & Baayen 2012). Further, many of the social predictors in my dataset are highly collinear (SEI, income, education, etc.), which complicates the analysis by mandating the construction of separate regression models for each predictor. Indicative of this complication is Chapter 6, in which I ran 16 separate mixed-effects regression models to address a single research question.

One powerful sub-component of the random-forest ensemble is the *variable importance analysis*, first designed by Breiman (2001) and improved by Strobl et al. (2008). The analysis rank-orders predictors – often highly collinear – without having their collinearity inflate importance (as is the case in linear regression models). Once the bagging ensemble has been completed, there are 'bags' of data that are not part of the final constructed 'forest'. These bags can then be used to calculate the importance of each variable, because the prediction accuracy of each bag can be compared with the prediction accuracy of bags that do not contain that specific predictor. This is calculated for all the predictors, rendering a variable-importance hierarchy. Tagliamonte and Baayen (2012) incorporate this tool to rank-order the predictors of *was/were* variation in York English (2012:160–162).

For the research question addressed in this section, [*RQ: To what degree is speech rhythm stylistically sensitive in Stockholm, and what inference can be made about the maturity of its associated variants?*], a variable importance analysis is particularly handy. This is because one would expect social predictors to decrease in importance as formality increases for a highly salient *marker-like* feature. Elseway, for low-salience *indicator-like* features, one would expect social predictors to maintain their importance or only decrease in importance by a mild degree.

Variable importance analysis

I used the *randomForestSRC* package in *R*, which is an improvement to the *party* package used in Tagliamonte and Baayen (2012). *randomForestSRC* commandeers all of the cores available in its hosting machine to run the analysis whereas the software package *party* limits itself to one core, rendering the computation too expensive. I ran three separate random forest analyses for each of the speech styles (CASUAL, READING, RADIO, respectively) using the following call:

> nPVIV ~ ACCENT_1_LONG + ACCENT_1_SHORT + ACCENT_2_LONG + ACCENT_2_SHORT + UNSTRESSED_LONG + UNSTRESSED_SHORT + CODA_R + PHRASE_FI-

Key	
ACCENT_1_LONG	contrast pairs that contain a lexical-pitch accent-1 long vowel (p. 130)
ACCENT_1_SHORT	contrast pairs that contain a lexical-pitch accent-1 short vowel (p. 130)
ACCENT_2_LONG	contrast pairs that contain a lexical-pitch accent-2 long vowel (p. 130)
ACCENT_2_SHORT	contrast pairs that contain a lexical-pitch accent-2 short vowel (p. 130)
UNSTRESSED_LONG	contrast pairs that contain an unstressed long vowel (p. 134)
UNSTRESSED_SHORT	contrast pairs that contain an unstressed short vowel (p. 134)
PHRASE_FINAL	contrast pairs that contain a vowel that is the nucleus of a phrase-final syllable (p. 134)
CODA_R	contrast pairs that contain a vowel that is succeeded by a syllable-coda /r/ (p. 136)
SPEECH_RATE	mean syllabic duration of the two syllables containing the contrast pair (p. 137)
LEXICAL_FREQ	lexical frequency of the word(s) in which the contrast pair occurs (p. 139)
AGE	age of the speaker in 2017 (p. 144)
RACIALIZATION	self-identification of the speaker as *invandrare* or *svensk* (p. 86, p. 144)
PC1_SOCIAL_CLASS	first principle component of a Principle Components Analysis of SEI, INCOME, EDUCATION, PARENTAL_SEI, PARENTAL_EDUCATION, and MOSAIC_TASTE (p. 89, p. 150)
SEI	speaker's occupational status according to Ganzeboom and Treiman (2003) (p. 145)
INCOME	speaker's monthly income in 1000SEK (p. 146)
EDUCATION	speaker's highest-attained educational level (p. 147)
PARENTAL_SEI	highest occupational status (Ganzeboom & Treiman 2003) of the speaker's parent(s) (p. 147)
PARENTAL_EDUCATION	average of the highest-attained educational level of the speaker's parent(s) (p. 147)
MOSAIC_TASTE	the status of the speaker's taste according to a 60-question survey derived from the *Mosaic Sweden* market segmentation schema (Experian 2013) (p. 148)
CLASSMATES_ FOREIGN_ BACKGROUND	weighted average of the percentage of classmates with a 'foreign background' of every school the participant attended, grades 7–12 (p. 165)
CLASSMATES_ WORKING_ CLASS	weighted average of the percentage of working-class classmates of every school the participant attended, grades 7–12 (p. 165)
LIFETIME_ NEIGHBORHOOD_ DIVERSITY	weighted average of the percentage of residents with a 'foreign background' of every neighborhood the participant has lived in (p. 165)
NETWORK_INVANDRARE_WC	percentage of close contacts, identified as *invandrare*, whose occupational SEI < 50 (p. 167)
NETWORK_SVENSK_WC	percentage of close contacts, identified as *svensk*, whose occupational SEI < 50 (p. 167)
NETWORK_SPEAKS_HOOD	percentage of close contacts who the participant says speaks *hood/ ghetto/suburban* (p. 167)

Table 8.4: Key for predictor codes used in statistical models.

NAL + LEXICAL_FREQ + SPEECH_RATE + AGE + RACIALIZATION + PC1_SOCIAL_CLASS + SEI + INCOME + EDUCATION + PARENTAL_SEI + PARENTAL_EDUCATION + MOSAIC_TASTE + CLASSMATES_FOREIGN_BACKGROUND + CLASSMATES_WORKING_CLASS + LIFETIME_NEIGHBORHOOD_DIVERSITY + NETWORK_SPEAKS_HOOD + NETWORK_SVENSK_WC + NETWORK_INVANDRARE_WC

Table 8.4 provides a key for the social predictors. Refer to Chapter 6 for the specifics on how they were operationalized.

Figure 8.5 contains a plot of each of the three variable importance outputs for nPVIV for each contextual style. On the left side is the entire plot, and it shows that internal factors dominate in variable importance, regardless of speech style. Age, however, shows no clear pattern in its effects on to contextual style. Rather, it appears to shadow the variable importance of three internal factors (unlabeled): *accent 1 long*, *accent 2 short*, *and speech rate*. This is something that will be investigated more closely in Chapter 9. Social factors generally decrease in importance from CASUAL to READING to RADIO.

The right side of Figure 8.5 shows an enlargement of the $y = [0, 0.01]$ portion with the internal factors removed. There, the individual social factors are labeled, and it becomes more clear that classic social factors, that is, the components of *Social class (PC1)* show the greatest decrease in importance. The network predictors, however, show much more variation. *Network svensk working-class*, *Network speaks hood*, and *classmates working-class* all show a decrease in importance. *Lifetime neighborhood diversity* and *Classmates foreign background* show stable variable importance, and *Network invandrare working-class* shows a monotonic climb in importance.

Returning to the question of social salience, the analysis at hand shows a monotonic decrease in importance for five of the six classic social predictors alongside their principle component. It shows a monotonic decrease in importance for three of the six social-network predictors. The evidence here points to rhythm, defined by nPVIV, as carrying sufficient social salience that it is adjusted in formal speech to the degree that the influence of social identity is attenuated.

4.3 Style as a predictor and separating formality from reading

The two analyses presented thus far present a somewhat tricky empirical challenge. Read-aloud and spontaneous speech have inherent mechanical differences. Therefore, it is possible that these differences are the reason why the social predictors have weakened effects in READING and RADIO. First, I explore the short literature on this topic. I then model style – and its interactions with the social predictors – as a predictor of nPVIV.

Spontaneous vs. read-aloud speech: A LabPhon perspective

The Laboratory Phonetics enterprise has found there to be general rhythmic differences between read-aloud and spontaneous speech. This presents a unique problem for the results presented here thus far. In other words, reading aloud puts mechanical constraints on production in ways that have nothing do with formality. In a recent large-scale analysis, Arvaniti (2012) examined speech rhythm in sentence reading, story reading, and spontaneous speech in English, German, Greek, Italian, Korean, and Spanish. She found that the nPVIV for spontaneous speech was significantly greater than in read-aloud speech for all languages, save German (2012:371). Mok and Lee (2008) compared rhythm between read-aloud and casual Korean and found that the spontaneous retelling of the *North Wind and Sun* had higher alternation than its formal reading.

It is unclear what would cause the above differences. Read-aloud speech has typically lower speed and higher average lexical frequency than spontaneous speech, which would pre-

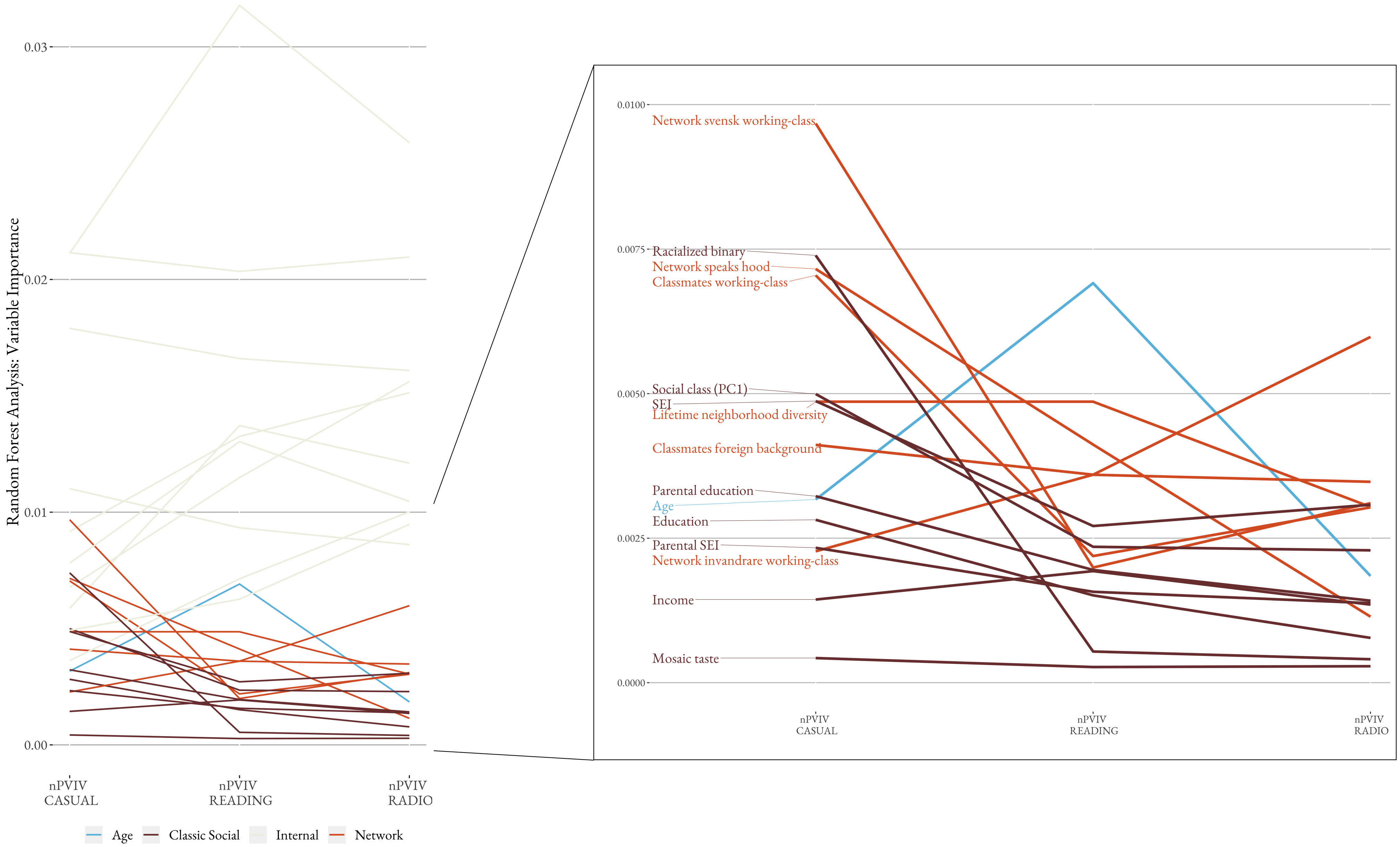


Figure 8.5: Comparative visualization of the three Random Forest variable importance analyses conducted for CASUAL, READING, and RADIO, each. The left plot shows the entire output with all predictors. The right plot is an enlarged cutout of $y = [0, 0.01]$ for just the non-internal predictors. The x-axis plots contextual style. The y-axis plots the *variable importance index*[a], developed by Breiman (2001), which designates that any value > 0 has influence. The blue line is *age*; white lines are *internal* predictors; brown lines are *classic social* predictors; orange lines are *network* predictors

[a] NB: The *variable importance index* constitutes its own measurement domain; in other words, it is not a proxy for, e.g., a p-value.

dict the *opposite* result of Arvaniti (2012) and Mok and Lee (2008). The literature has shown that slower speech is correlated with higher nPVIV and vice versa (see review on p. 137). Returning to my dataset, let us examine speech rate. Mean speech rate among speakers is 148 ms for casual and 164 ms and 165 ms for reading and radio, respectively (p. 219). In other words, the shift from casual to reading (or casual to radio) ought to result in a higher nPVIV across the board.

Lexical frequency is also highly varied (p. 219) due to the rather unusual topic of the reading passage (a boy and his grandfather at a circus). The mean lexical frequency in the Gothenburg Spoken Corpus for the words produced in the casual portion of the corpus is 7 415. For the reading and radio portions it is 2 333 and 2 301, respectively. Recalling the review in Section 2.2, low lexical frequency has been shown to attenuate vowel reduction (Pluymaekers et al. 2005; Zhao & Jurafsky 2009). Therefore – as could be the case for speech rate – low lexical frequency might also increase nPVIV.

Another possibility is mid-phrase lengthening in the case of changing lines. For example, take the following excerpt from the reading passage:

> På ängen hade man rest ett enormt cirkustält. Vi var först i kön och fick platser
> närmast manegen. När ljuset mörknade spratt jag till av förväntan. Cirkusorkestern
> spelade upp och föreställningen började klockan 20.00 på pricken

It is possible that readers will lengthen the last syllable of *platser* or *cirkusorkestern* as they search for the next line.

In the three possibilities I cite above, one would expect reading aloud to have a uniform increasing effect on nPVIV. By contrast, the literature reviewed above would predict reading aloud to have a uniform decreasing effect. In either case, we would expect these directional effects to occur independent of social identity.

Including style as a predictor in the regression model

The prior analyses have modeled each contextual style in separate models, which means that the main effect of style – and whether it is significant or not – has gone unmeasured. I therefore constructed a mixed effects linear regression model that includes contextual style as a predictor with the following call:

> nPVIV ~ accent_1_long + accent_1_short + accent_2_long + accent_2_short + unstressed_long + unstressed_short + coda_r + phrase_final + log(lexical_freq) + log(speech_rate) + age + racialization*pc1_social_class*style + (1|speaker) + (1|vowel)

The model has the social predictors age, class, racialization, the interaction of class and racialization, and the interaction of contextual style with the aforementioned. If there are internal mechanical properties to read-aloud speech that effect rhythm – independent of any social factors – then there ought to be a unitary predicted response in nPVIV.

	Model 1 *Contextual style*	
Response variable	nPVIV (all styles)	
(Intercept)	0.3 (8.0)	
Internal predictors		VIF
accent_1_long·*yes*	24.2 (0.9)***	1.3
accent_1_short·*yes*	14.6 (0.9)***	1.1
accent_2_long·*yes*	20.8 (1.1)***	1.2
accent_2_short·*yes*	10.8 (0.9)***	1.1
unstressed_long·*yes*	8.4 (0.8)***	1.4
unstressed_short·*yes*	17.6 (0.9)***	1.2
coda_r·*yes*	8.1 (1.2)***	1.0
phrase_final·*yes*	7.5 (0.8)***	1.0
log(lexical_freq)	−0.2 (0.1)*	1.1
log(speech_rate)	6.1 (0.7)***	1.1
Style alone predictors		VIF
style·*reading*	−17.02 (2.0)***	3.3
style·*radio*	−13.93 (2.0)***	3.3
Social predictors		VIF
age	0.63 (0.2)**	1.2
racialization·*invandrare*	−21.92 (4.6)***	2.2
class	−0.18 (0.1)**	1.8
racialization·*invandrare*:class	0.26 (0.1)**	2.1
Style interaction predictors		VIF
style·*reading*:racialization·*invandrare*	16.97 (2.3)***	3.1
style·*reading*:class	0.24 (0.0)***	2.9
style·*reading*:racialization·*invandrare*:class	−0.23 (0.0)***	2.4
style·*radio*:racialization·*invandrare*	16.79 (2.3)***	3.1
style·*radio*:class	0.21 (0.0)***	2.9
style·*radio*:racialization·*invandrare*:class	−0.23 (0.0)***	2.4
aic	415 415.4	
bic	415 639.0	
Log Likelihood	−207 681.7	
Num. obs.	40 277	
Num. groups: vowel	3 334	
Num. groups: speaker	36	
Var: vowel (Intercept)	176.8	
Var: speaker (Intercept)	35.3	
Var: Residual	1 689.4	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Table 8.5: Mixed-effects linear regression model that includes age, class, racialization and contextual as predictors of nPVIV. For categorical predictors, the reference category is in italics. Coefficients are indicated in the center column, standard errors in the parentheses, and variance inflation factors (VIF) to the right.

Following, for example, Arvaniti's (2012) finding, speech rhythm should unidirectionally decrease for all speakers. Or, if we were to follow my three examples, speech rhythm should unidirectionally increase for all speakers.

Table 8.5 contains the output of the regression model. casual speech is the reference style, so the main effects of the social predictors resemble the earlier models (regression outputs on pp. 157, 173, and 221). *Invandrare* identity is predicted to reduce nPVIV, lower social class combined with *svensk* identity is predicted to increase nPVIV, and lower social class combined with *invandrare* identity is predicted to decrease nPVIV even further for casual speech.

A shift from casual to reading has a negative main effect, in line with Arvaniti's (2012) finding. However, it predicts an overall increase for *invandrare*, especially *invandrare* with a lower class index. A shift from casual to reading predicts an overall decrease for *svensk*, especially *svensk* with a lower class index. The shift from casual to radio has nearly identical coefficients. The main effects for style and its interactions are all statistically significant.

The model shows that style has a negative main effect on the data and that it is heavily influenced by social factors. These social factors interact with style such that a contextual style shift will not render a unidirectional change in nPVIV across the whole population. Rather, nPVIV moves in different directions, depending on the speakers' background. This implies that more is at play than inherent mechanical differences between spontaneous and read-aloud speech. It implies that contextual style, and its associated formality, is the more important governing factor.

5 Chapter summary

This analysis sought to address the following research question: *RQ: To what degree is speech rhythm stylistically sensitive in Stockholm, and what inference can be made about the maturity of its associated variants?*. The findings are provided below.

1. Working-class *invandrare* speakers produce *staccato* speech in CASUAL, and it becomes *less staccato* in READING and RADIO.

2. Working-class *svensk* speakers produce *non-staccato* speech in CASUAL, and it becomes *more staccato* in READING and RADIO.

3. Upper-class speakers, *svensk* and invandrare, produce an intermediate alternation pattern in CASUAL, and it changes very little in READING and RADIO.

4. Working-class speakers, *svensk* and invandrare, produce a rhythmic pattern in READING and RADIO that resembles the CASUAL pattern of upper-class speakers.

5. Younger speakers, regardless of social background, are predicted to have more staccato speech than older.

The above results derive from the cumulative assessment of three analyses.

First, the predictive power of racialization, class, and their interactions decline monotonically from CASUAL to READING to RADIO in mixed-effects regression models. The predictive power of internal factors and age remains strong and stabile.

Second, I validated the first analysis by assessing the predictive power of all social-class artifacts and all social-network qualities on nPVIV. Five of the six social-class predictors decline in variable importance as formality increases. Three of six social-network effects also show decreases.

Third, contextual style and its interaction with the social predictors were modeled in a single regression analysis. Style had a significant negative main effect, which implies that a switch from spontaneous to read-aloud speech results in lower rhythmic alternation by virtue of some properties inherent to reading aloud. In addition to this main effect, age, class, racialization, the interaction of class with racialization, and the interaction of style with class and racialization all had significant effects. These interaction effects strongly coerce the main effect of style on nPVIV, rendering no unidirectional shift in nPVIV across all social groups.

Age is not the focus of this investigation, but the findings warrant bringing it up here. This is because it persists as a significant effect in nearly every model run, regardless of contextual style. The effect is mild and indicates between a 0.5 and 0.7 increase in nPVIV for every year older the speaker has. It is an effect that applies to the full speaker dataset and not just the targeted *invandrare* working-class speakers. The next chapter will deal with age in more detail, including its interaction with contextual style and the phonology of rhythm.

The findings here imply that both the staccato rhythm of the *invandrare* working class and the non-staccato rhythm of the *svensk* working class have some degree of social salience and, therefore, some degree of maturity in the speech community. In the next chapter, their respective developments will be examined more closely in the context of apparent time.

Chapter 9

RHYTHM AND CHANGE IN STOCKHOLM

In Chapter 8, I examined intraspeaker variation to make informed hypotheses about change. The stylistic sensitivity of nPVIV among the working classes, both *invandrare* and *svensk*, implies that rhythm is a stratified variable whose staccato and non-staccato variants have circulated for some period of time. This argument is strengthened by the literature that has alluded to staccato rhythm since the late 1980s. The current chapter is a continuation of chapter 8. The main part of the analysis will rely on age as a signifier of *change in apparent time* (Labov 1994:43–72), and the focus will be on: (1) whether stylistic sensitivity changes over time; (2) the change point at age 33 for speech rhythm and how this correlates with macrosocial changes in Stockholm; (3) how that same change point correlates with different phonological transformations among young speakers and different phonological transformations among older speakers; (4) whether the overall lowering of nPVIV across Stockholm is connected to its multiethnolect.

1 Research question

A single broad research question shapes the current analysis:

> *RQ: What is the apparent-time stratification of rhythm in Stockholm, and what does this imply about its evolution?*

Four analyses will be conducted to address the above question. Each section will contain an analysis.

In the first analysis, I wish to understand whether stylistic sensitivity changes over time. Now that low nPVIV has been established as a feature among the *invandrare* working class, can an evolution from indicator to marker be seen? In other words, are younger speakers more stylistically sensitive than older speakers? Now that high nPVIV has been established as a feature among the *svensk* working class, can an evolution from indicator to marker be seen there? These questions are addressed by building on the statistical model in Chapter 8 and modeling age as an interaction effect with style.

In the second analysis, I test my impression that older speakers of multiethnolect are doing something different to rhythm than younger speakers in casual speech. I re-examine the age trends within each of the three contextual styles and identify a breakpoint at age 33 within each for nPVIV. I then discuss the macrosocial correlates that are contemporaneous with this breakpoint.

In the third section, I build on the previous analysis. Here, I only examine the CASUAL dataset and ignore the more formal styles. I use a conditional inference tree to find critical break points in the phonology as it relates to age, racialization, and social class. I trace the progression of staccato rhythm across empirically-established cohorts – in other words, clusters of speakers established by the model. I then tease apart which phonological transformations are unique to each of these cohorts. From there, I establish that there are seven social dialects operating in the dataset as it relates to statistically-significant clusters of speech-rhythm variation.

In the fourth and final section, I pilot an exploratory analysis that seeks to address the reason why younger higher-class speakers have such low nPVIV values. Could it be related to the diffusion of multiethnolectal features? I investigate weak ties within social networks and the effects of these ties on nPVIV in order to establish an informed hypothesis for eventual expansion and follow-up.

	Model from Table 8.5 (p. 230)		New Model *Style and age interaction*	
Response variable	nPVIV (all styles)		nPVIV (all styles)	
(Intercept)	0.3 (8.0)		−2.7 (8.1)	
Internal predictors		VIF		VIF
accent_1_long·*yes*	24.2 (0.9)***	1.3	24.2 (0.9)***	1.3
accent_1_short·*yes*	14.6 (0.9)***	1.1	14.6 (0.9)***	1.1
accent_2_long·*yes*	20.8 (1.1)***	1.2	20.8 (1.1)***	1.2
accent_2_short·*yes*	10.8 (0.9)***	1.1	10.8 (0.9)***	1.1
unstressed_long·*yes*	8.4 (0.8)***	1.4	8.4 (0.8)***	1.4
unstressed_short·*yes*	17.6 (0.9)***	1.2	17.6 (0.9)***	1.2
coda_r·*yes*	8.1 (1.2)***	1.0	8.1 (1.2)***	1.0
phrase_final·*yes*	7.5 (0.8)***	1.0	7.5 (0.8)***	1.0
log(lexical_freq)	−0.2 (0.1)*	1.1	−0.2 (0.1)*	1.1
log(speech_rate)	6.1 (0.7)***	1.1	6.1 (0.7)***	1.1
Style alone predictors		VIF		VIF
style·*radio*	−13.9 (2.0)***	3.3	−7.69 (3.6)*	6.1
style·*reading*	−17.0 (2.0)***	3.3	−11.24 (3.6)**	6.1
Social predictors		VIF		VIF
age	0.6 (0.2)**	1.2	0.73 (0.2)***	1.3
racialization·*invandrare*	−21.9 (4.6)***	2.2	−21.88 (4.6)***	2.2
class	−0.2 (0.1)**	1.8	−0.19 (0.1)**	1.8
racialization·*invandrare*:class	0.3 (0.1)**	2.1	0.27 (0.1)***	2.1
Style interaction predictors		VIF		VIF
age:style·*reading*	-		−0.21 (0.1)$^{\circ}$	5.9
age:style·*radio*	-		−0.23 (0.1)*	5.9
style·*reading*:racialization·*invandrare*	17.0 (2.3)***	3.1	16.93 (2.3)***	3.1
style·*reading*:class	0.2 (0.0)***	2.9	0.26 (0.0)***	3.1
style·*reading*:racialization·*invandrare*:class	−0.2 (0.0)***	2.4	−0.24 (0.0)***	2.4
style·*radio*:racialization·*invandrare*	16.8 (2.3)***	3.1	16.73 (2.3)***	3.1
style·*radio*:class	0.2 (0.0)***	2.9	0.24 (0.0)***	3.1
style·*radio*:racialization·*invandrare*:class	−0.2 (0.0)***	2.4	−0.25 (0.0)***	2.4
aic	415 415.4		415 418.6	
bic	415 639.0		415 659.5	
Log Likelihood	−207 681.7		−207 681.3	
Num. obs.	40 277		40 277	
Num. groups: vowel	3 334		3 334	
Num. groups: speaker	36		36	
Var: vowel (Intercept)	176.8		176.6	
Var: speaker (Intercept)	35.3		35.3	
Var: Residual	1 689.4		1 689.3	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Table 9.1: Mixed-effects linear regression models with nPVIV as the response variable. All models contain the same internal predictors. The left model is from Table 8.5 (p. 230). In it, the external predictors are age, racialization, social class, the interaction of racialization with social class, style, the interaction of style with racialization, and the interaction of style with the interaction of racialization with social class. In the New Model to the right, the same external predictors are present with one additional external predictor: the interaction of age with style. For categorical predictors, the reference category is in italics. Coefficients are indicated in the center column, standard errors in the parentheses, and variance inflation factors (VIF) to the right. The external predictors in New Model are rounded to two decimal points to show their finer differences. The remaining are rounded to one decimal point to save space.

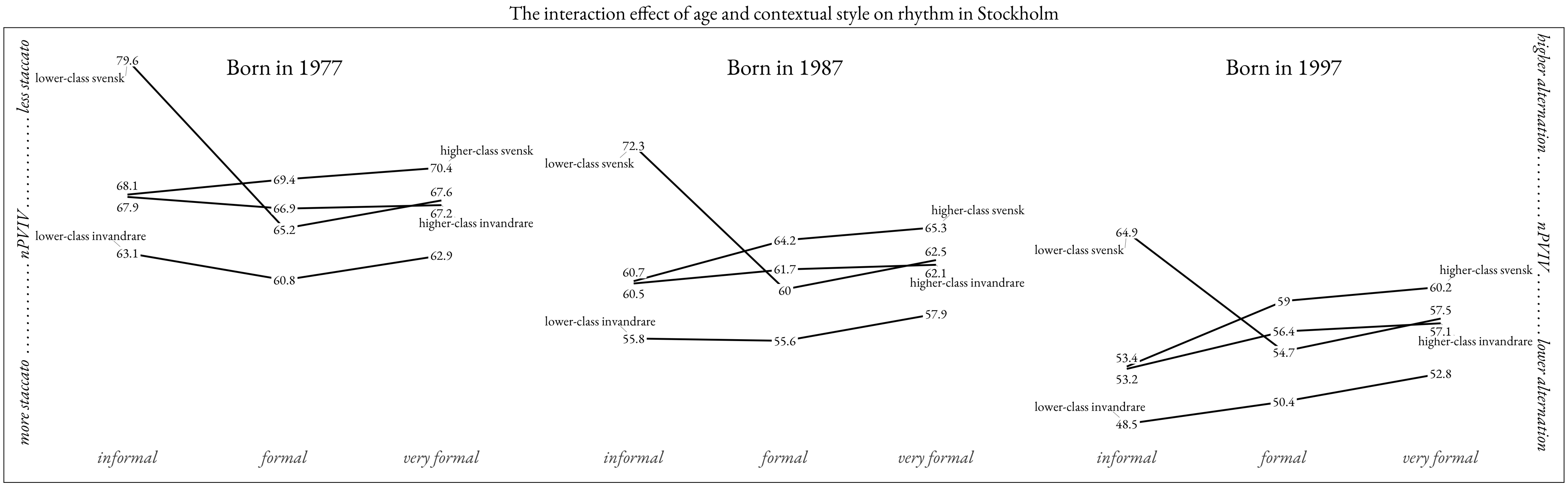


Figure 9.1: The interaction effect of age and contextual style on rhythm: A simulation built from the coefficients in Table 9.1 regression Model 1 (p. 235) for four hypothetical speakers in three contextual styles from three cohorts. Hypothetical speakers are lower-class *svensk*, lower-class *invandrare*, higher-class *svensk*, higher-class *invandrare*. NB that lower class is modeled with a class index of 20 and that higher class is modeled with a class index of 80. 'Lower' and 'higher' refer to relative placement in this dataset. Contextual styles are *informal*, *formal* and *very formal*. Cohorts are born in 1977 (age 40), born in 1987 (age 30), and born in 1997 (age 20).

2 Style and age in interaction

In Chapter 8, age had a consistent main effect in all models run. In this analysis, I wish to understand the development of rhythmic stratification in apparent time while also identifying changes to stylistic sensitivity in apparent time. Therefore, I took the same model from Table 8.5 (p. 230) and added style and age as interactions:

> nPVIV ~ ACCENT_1_LONG + ACCENT_1_SHORT + ACCENT_2_LONG + ACCENT_2_SHORT + UNSTRESSED_LONG + UNSTRESSED_SHORT + CODA_R + PHRASE_FINAL + *log*(LEXICAL_FREQ) + *log*(SPEECH_RATE) + AGE + AGE:STYLE + STYLE* RACIALIZATION*PCI_SOCIAL_CLASS + (1|SPEAKER) + (1|VOWEL)

Table 9.1 contains the output from this model alongside a reproduction of the model from Table 8.5 (p. 230) in order to facilitate a comparison. It shows that age and style interact significantly in the shift from CASUAL to RADIO and near-significantly in the shift from CASUAL to READING. Meanwhile, other model indicators remain robust: the AIC and BIC penalties are only slightly greater, the effects of the other predictors remain stable, and the variance inflation factors are still less than 10.

Interpreting the model, however, is tedious if one is to rely solely on an examination of the coefficients. I therefore offer a case study built on the coefficients from this model. This is shown in Figure 9.1. Four speakers from opposite ends of the class and racialization spectrum are modeled: lower-class and higher-class *svensk* and lower-class and higher-class *invandrare*. NB that 'lower' and 'higher' reflect relative placement within this dataset; particularly in the case of 'higher', this does not necessarily mean upper class and could be more reflective of middle or upper-middle class (see p. 258 for a more thorough discussion). A class index of 20 is used for lower class, and a class index of 80 is used for higher class. Age is modeled at three intervals: 40 (born in 1977), 30 (born in 1987), and 20 (born in 1997).

The trends in Figure 9.1 are particularly informative to the research question. Overall, nPVIV decreases with time, independent of other social factors. However, the interaction with style enriches this picture. Among lower-class *invandrare* speakers, staccato rhythm is a consistent variant that continues to become *more* divergent from the mainstream in informal speech between the 1977 and 1997 cohorts while also becoming much more stylistically sensitive. This implies movement from *indicator* to *marker* status as well as a *hypercorrection from below*. That is to say "successive generations carry the variable further along the process of change" (Labov 1972a:178). As the variant becomes more divergent in the 1987 and 1997 cohorts, it also appears to become more socially salient. Among lower-class *svensk* speakers, high-alternation rhythm is a consistent vernacular variant that continues to maintain its divergence from the mainstream in informal speech between the 1977 and 1997 cohorts while also maintaining a high degree of stylistic sensitivity. This implies that their non-staccato variant is more established than its staccato counterpart, perhaps a legacy feature of Stockholm's Industrial-era Södersnack (see review on pp. 47–48).

3 A change point in nPVIV corresponds with certain historical events

The staccato rhythm identified by Kotsinas is not necessarily the same staccato rhythm heard in Stockholm's suburbs today. This is implied by the case study in Figure 9.1 and is consistent with an impression I long have had that Kotsinas' generation of speakers produced a much milder form of staccato rhythm than the younger speakers in this study. My colleague Dr. Kari Fraurud agreed in a personal conversation in 2014. She observed that the speech in the more recent SUF[1] and SALAM[2] studies was more 'foreign-sounding' than the speech from Kotsinas' era. In that same conversation, she pointed out that her assessment was based on an early video recording of Rinkeby Swedish made in 1985 (Sveriges Television 1985) and could therefore not be written off as age grading. In a separate project, I have begun collecting media recordings of Rinkeby Swedish from as early as 1979 (e.g., Sveriges Television 1979). In these early recordings, the Swedish also sounds more standard-like than that of any of speakers in this study, including the older ones.

In some respects, the difference in impression mirrors the dynamic of London Jamaican versus Multicultural London English (Cheshire, Kerswill, Fox, & Torgersen 2011). Should one hear recordings of London Jamaican today, one gets the impression that it sounds quite Cockney in comparison to the impression one gets from MLE. Similarly, to modern ears, 1980s Rinkeby Swedish sounds remarkably Södersnack-like.

In Chapter 8, I identified a positive effect of age on nPVIV and interpreted this as change in apparent time. What the regression model failed to reveal – and this is one shortcoming of linear regressions in general – was a critical breakpoint in the age data. Figure 9.2 shows a reproduction of Figure 8.4 (p. 223) with its linear smoothers. Below it, Figure 9.3 shows the same data, albeit with locally estimated scatterplot smoothers (LOESS, Baayen 1996; Cleveland 1979). The smoothers reveal a change-point in all styles at approximately age 33 whereby the *invandrare* and *svensk* speakers start to diverge dramatically.

In the subsequent paragraphs, I will present historical data that documents macrosocial change points that occurred in proximity to when these speakers were children. I draw a connection between these historical events and the shift in speech rhythm.

3.1 A spike in migration

As I reviewed on pages 20–24, Sweden underwent a second major migration surge in the mid 1990s. Figure 9.4 shows a graph of Sweden's population change from 1875 to 2018 (Statistics Sweden 2017). In it, three spikes in migration can be seen. The first occurs between 1968 and 1973 and was due to Sweden's postwar guest-worker program. The second spike occurs

[1] Källström and Lindberg (2011)
[2] Bijvoet and Fraurud (2012)

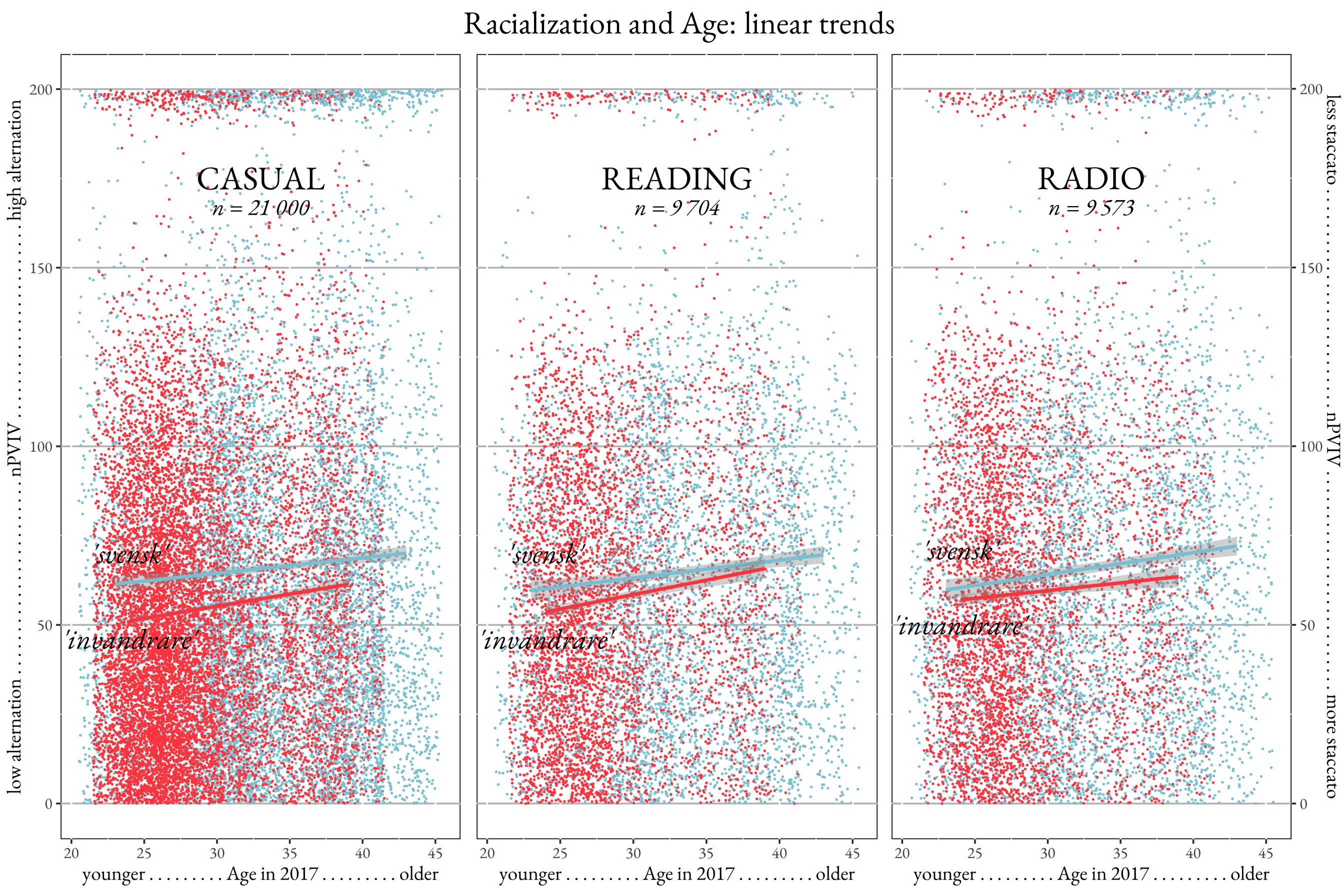


Figure 9.2: Reproduction of Figure 8.4: Scatterplots of nPVIV in CASUAL, READING, and RADIO according to age in 2017. *Svensk* speakers are in blue; *invandrare* speakers in red. Least-squares regression lines are plotted with confidence-interval halos.

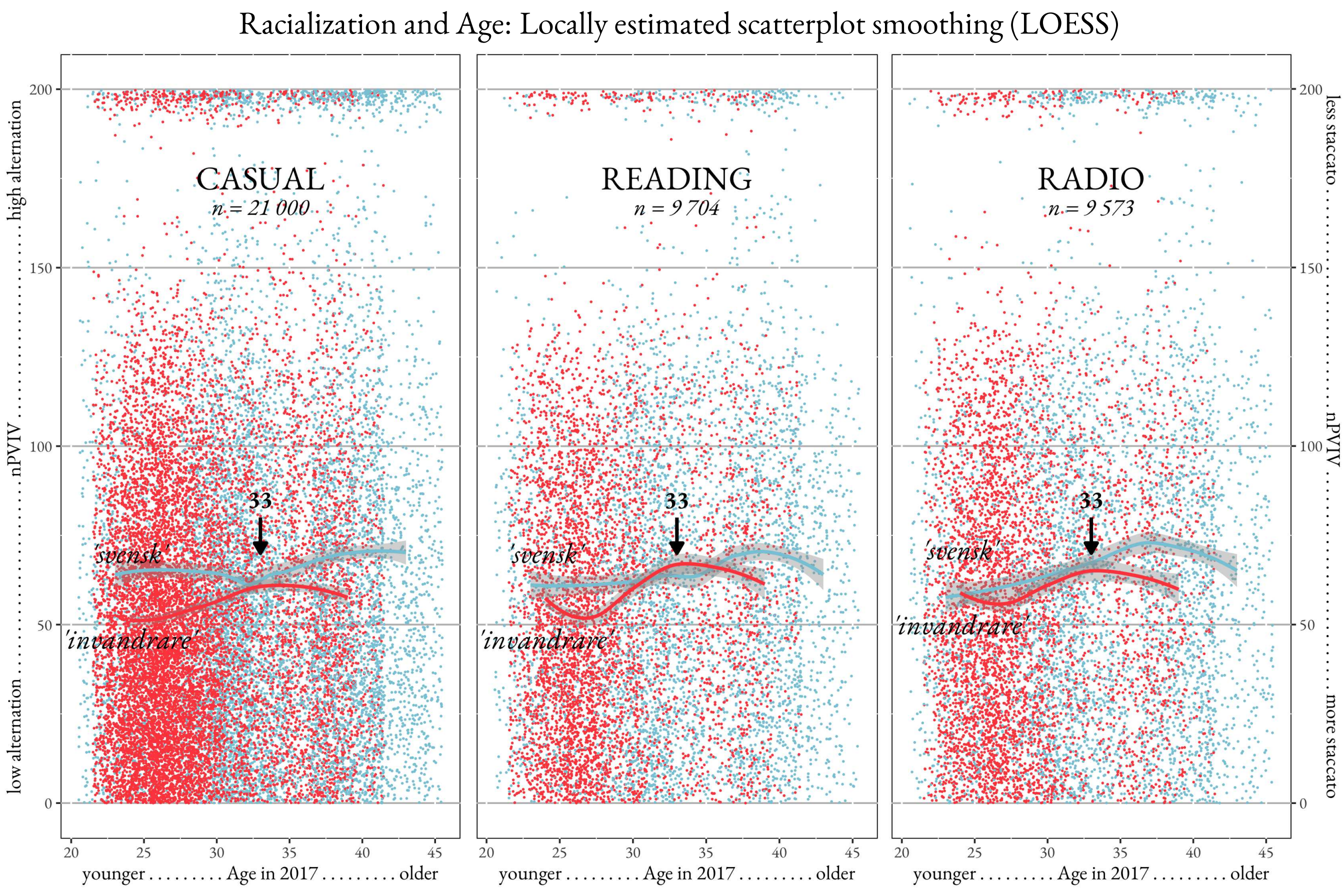


Figure 9.3: Scatterplots of nPVIV in CASUAL, READING, and RADIO according to age in 2017. *Svensk* speakers are in blue; *invandrare* speakers in red. Locally estimated scatterplot smoothing (LOESS) lines are plotted with confidence-interval halos.

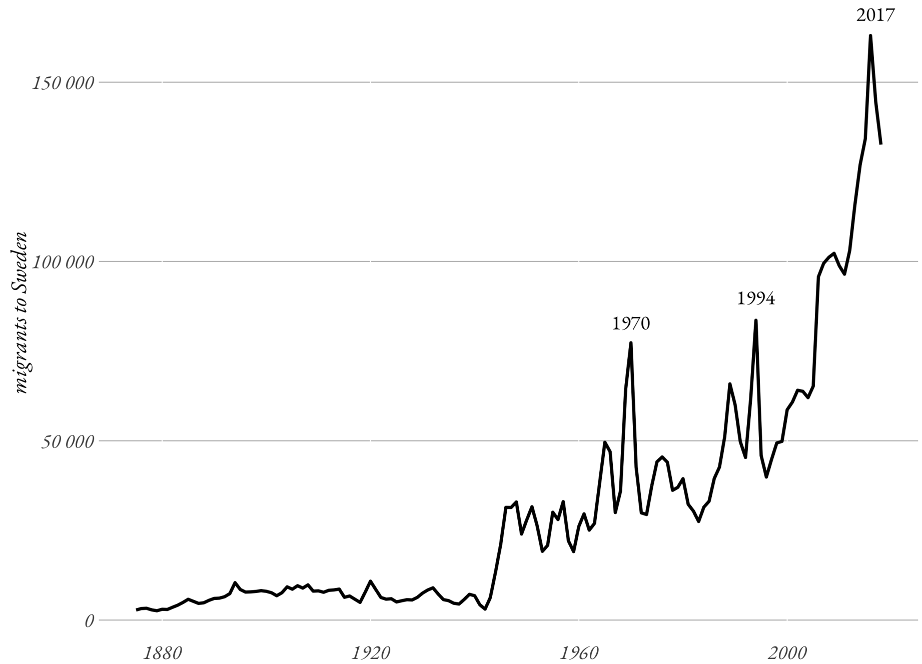


Figure 9.4: Immigrants to Sweden by year, 1875–2018 (Statistics Sweden 2017). Three immigration spikes are labeled: 1970 and surrounding years, 1994 and surrounding years, 2017 and surrounding years.

between 1987 and 1996 and was due to a number of simultaneous global conflicts. According to Statistics Sweden (2017), these were migrants "fleeing war and dictatorship in, for example, Lebanon, Iran and Chile. When Yugoslavia split into smaller parts in a civil war during the 1990s, many sought out Sweden for sanctuary" (Statistics Sweden 2017, my translation). The third spike occurs between 2015 and 2018 and is due to the war in Syria, during which Sweden accepted a record 574 336 migrants, increasing its total national population by nearly six percent.

The first spike coincides with the birth of a cohort that would have hit adolescence in the early 1980s – the very group that 'founded' Rinkeby Swedish and was the target of Kotsinas' investigation. At the time of the present study (2017), this cohort would be between the ages of 30 and 48. The second spike coincides with the birth of a cohort that would have hit adolescence between 1999 and 2008. This is the group that was the target of the SUF (Källström & Lindberg 2011) and SALAM (Bijvoet & Fraurud 2012) studies that emerged at the end of the 2000s. At the time of the present study (2017), this cohort would be approximately between the ages of 24 and 33.

3.2 A change point in school segregation

Not only would this second migration wave have led to a critical number of foreign-born adolescents in the late 1990s and early 2000s, it would have coincided with an unprecedented de-

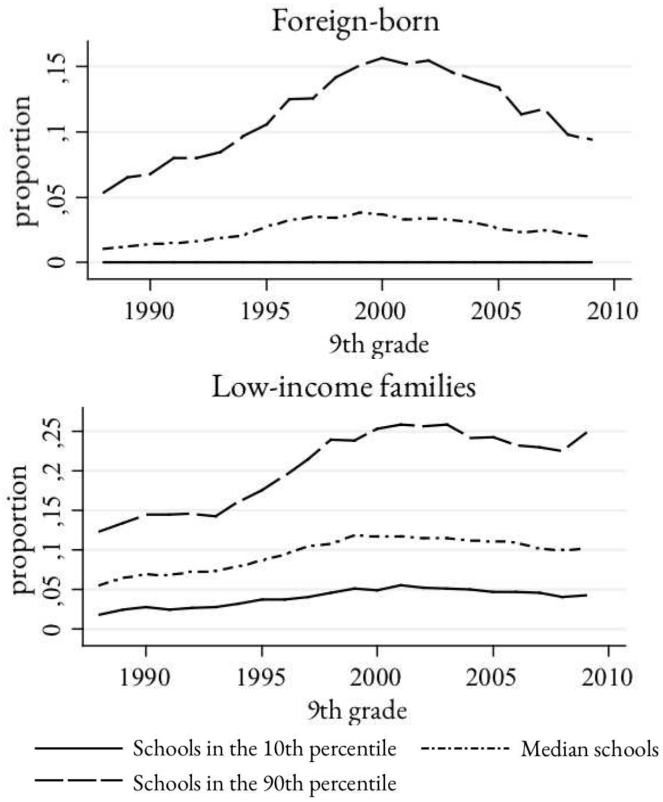


Figure 9.5: School segregation metrics for all primary schools in Sweden, 1988–2009 (reproduced and translated from Holmlund et al. 2014:88–90). *x*-axis is the year of the measurement. *y*-axis is the proportion of students in each school type as measured in the ninth grade.

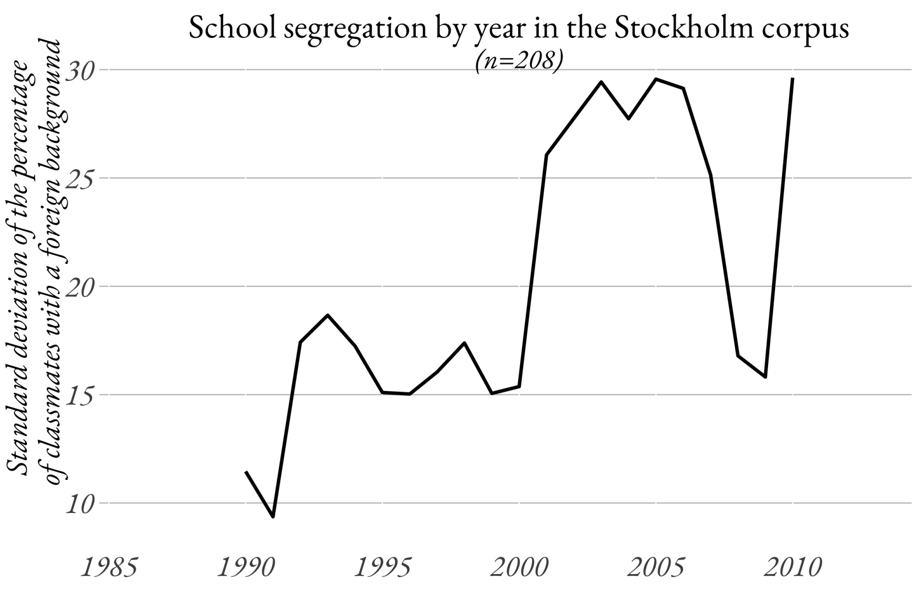


Figure 9.6: School segregation metrics for the 208 school-attendance years attended by the 36 speakers in the current Stockholm dataset, years 1990–2010, from grades 6 through 12 (*högstadium* and *gymnasium*). *x*-axis is the year of speaker attendance in school. *y*-axis is the standard deviation of the percentage of students of a 'foreign background' for the schools listed by participants for that given year. 'Foreign background' is a statistical category defined by Statistics Sweden (2002) as having both parents born abroad. Year- and school-specific demographic data were procured in 2018 from The Swedish National Agency for Education (*Skolverket*).

gree of school segregation. As I discussed on pages 29–31, the inception of Rinkeby Swedish coincided with stark neoliberal trends in Sweden. This includes growing inequality, residential segregation, and school segregation. Figure 9.5 contains a reproduction of two graphs from Holmlund et al.'s (2014) investigation of the relationship between school segregation and the 1992 charter-school reform and the 2001 teacher reform. It shows a peak at years 2000–2004 for segregation in terms of ethnic and socioeconomic lines. The top chart shows that the percentage of foreign-born students in the worst-performing schools (10th percentile) across Sweden climbed from 5% in 1988 to over 15% in 2000. The bottom chart shows that the percentage of low-income families in the worst-performing schools (10th percentile) across Sweden climbed from 12% in 1988 to 25% in 2000.

In my own dataset, the trends on ethnic segregation in schools mirror Holmlund et al.'s (2014) findings. Figure 9.6 contains a plot that illustrates the segregation trends for the 208 school-attendance years attended by the 36 speakers in the Stockholm dataset between 1990 and 2010. Only grades 6 through 12 (*högstadium* and *gymnasium*) were included because lower grades did not have reliable data (see p. 165). Segregation is operationalized by taking the standard deviation of the percentage of students with a 'foreign background' for each year observed. 'Foreign background' is a statistical category defined by Statistics Sweden (2002) as having both parents born abroad. Year-specific and school-specific demographic data were procured in 2018 from The Swedish National Agency for Education (*Skolverket*, see p. 165). Years 1985–1989 and 2011–2013 were excluded due to low representation.

As is the case for Holmlund et al.'s (2014) data, the Stockholm dataset in Figure 9.6 shows a similar change-point between year 2000 and 2001. Between 1990 and 2000, the median standard deviation of the percentage of students of foreign background is 15.3. Between 2001 and 2010, it is 27.7.

3.3 Year 2000: The largest income gap in modern Swedish history

In the very same year that school segregation shifted, the disposable-income gap peaked among the highest earners in Sweden and troughed among the lowest earners. Figure 9.7 contains the same graph that I reviewed on pages 31–35. In year 2000, disposable income reached a historical high for the top 10% of income earners and a historical low for the bottom 10% and 20% of income earners. Naturally, income inequality alone does not motivate stylistic practice, but byproducts of inequality do. One such byproduct is a decrease in social trust, which can be felt by adolescents and motivate practices of enclosure.

Gustavsson and Jordahl (2008) conducted a detailed analysis of social trust and income inequality using 1994 and 1998 Swedish election data ($n = 2\,792$). They found significant negative correlations between the 90/10-quotient for disposable income and social trust. Swedish counties with higher income inequality had lower social trust and vice versa.

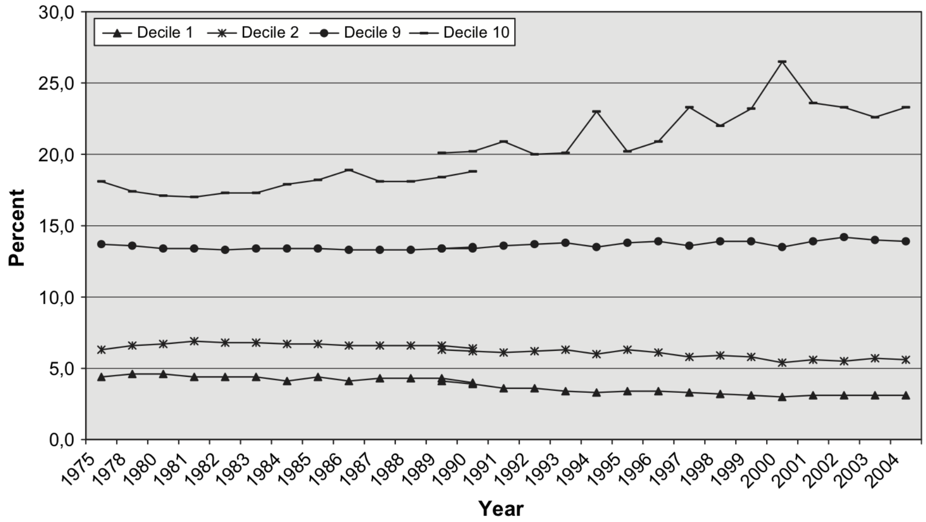


Figure 9.7: Income share (disposable income) for deciles 1, 2, 9 and 10 (1975–2004), reproduced from (Gustafsson & Jansson 2008:966)

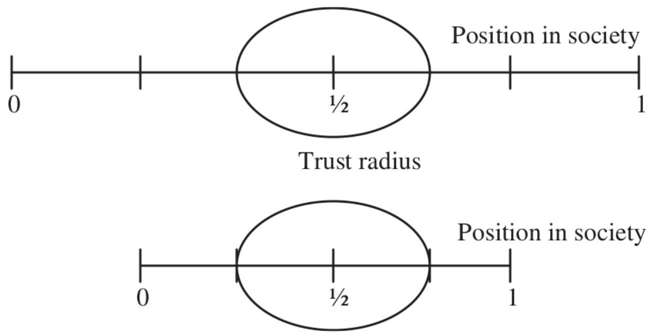


Figure 9.8: Fractionalization and trust radii. Reproduced from Bjørnskov (2008:273). The top line represents an individual's trust radius within a highly stratified society. The bottom line represents that same individual within a less-stratified society.

Bjørnskov (2008) found in his analysis of social-trust data of 100 countries that income inequality had a significant inverse relationship to social trust[3]. He contextualizes this phenomenon within the framework of *fractionalization* and Fukuyama's (1995) *trust radius*, reproduced in Figure 9.8. According to Uslaner (2002), a trust radius is equivalent to a moral community and that "the rich and the poor have little reason to believe that they share common values and thus might well be wary of each others' motives" (2002:181). Bjørnskov (2008) argues that the distribution of trust radii are equal in every society. In other words, there is a relatively fixed number of individuals who tolerate x number of stylistic differences before they distrust. There is a relatively fixed number of individuals who tolerate $x+1$ number of stylistic differences before they distrust, and so on, and so on. When a society becomes

[3]NB that this was negated by smaller countries. Sweden, however, does not count as small because the median population number in his dataset was 10 million (Bjørnskov 2008:278).

increasingly fractionalized, depicted in Figure 9.8, it results in more distrust events because the range of stylistic practices is stretched while the trust radius of each member remains the same (the model ignores individual adaptation). This results in overall decreased levels of trust. The connection to language change is that lower trust may motivate stylistic moves of enclosure. Among working-class adolescents, one such move is the formation of linguistic innovations. Another such move, related to the former, is *hypercorrections from below* (Labov 1972a), a term I will review in the next section.

In isolation, I would have written off the income trend as coincidence. However, when assessed alongside the other historical events reviewed above, year 2000 begins to look more and more like a transformative year.

3.4 Metalinguistic commentary about 'old Rinkeby Swedish'

In my ethnographic work I witnessed metalinguistic commentary concerning the generational split in Stockholm's multiethnolect. I have a recording in which three participants from Tensta – Mezdar, Max, Shorty – discuss how the 'older' men in their community speak. One such 'older' man is Solomon from this study. Mezdar, Max, and Shorty are all in their late twenties, and Solomon is in his late thirties (refer to participant list on page 140). An older man by the name of Berra is also mentioned. He did not participate in this study, but I have seen him around in Tensta and estimate him to be in his late thirties or early forties. The topic emerged at the end of a casual-speech recording in which I was not present:

Shorty: And it feels like Suburban Swedish is more full of love.

Mezdar: Yeah {yawn}

Shorty: I mean, it's more, you know, familiar.

Mezdar: You can show more love, bro.

Shorty: Everybody's brothers, everybody's sisters, everybody's... you feel me?

Mezdar: Yeah, it's not cold.

Shorty: Nah

Mezdar: But bro, have you thought about something? Those who are older, in Solomon's generation, for example.

Shorty: Yeah, Yeah say ((ninety-seveners, eighters[4]))

Mezdar: Solomon, eh, that guy that owns Café Mynta.

Shorty: Berra!

Max: He speaks clean Swedish.

Shorty: Yeah

Mezdar: Not... Their Swedish, it's not, eh, clean. It's a little... [*intonation implying something is off*]

Shorty: Yeah yeah yeah

Mezdar: It's less than...

Max: It's more clear and it's, um...

Mezdar: More clear, and they have...

[4] This appears to be an allusion to graduating class.

Max: There isn't so much sick slang, so to speak.

Mezdar: No, they sound more, like, Swedish.

Max: There's no, yeah, there's {laugh} no hint at all [presumably about their status as second-generation immigrants]

Shorty: Exactly, exactly.

Mezdar: Compared to us, it's totally clean, but compared to a Swede, you can tell that this dude...

Shorty: He's a little...

Mezdar: He's been around the block a time or two {laugh}, you know what I mean.

Shorty: Yeah yeah yeah yeah yeah {laugh} yeah, he's a little bit hood.

Mezdar: Yeah and then us, it's a little bit *more*. There's come new words, new invented words.

Shorty: yeah yeah yeah yeah yeah yeah yeah

Mezdar: Using words... we have words from other languages.

Shorty: Putting together two languages.

Mezdar: But bro, now those younger than us, bro, I don't even know what they're saying.

Shorty: We can't keep up![i]

What is particularly insightful from this discussion is the observation that older speakers from Tensta sound 'totally clean' compared to the younger speakers but still not completely Swedish-sounding compared to a 'real Swede'. Assessed in isolation, this discussion looks like a lay assessment of age grading. However, within the context of the findings on speech rhythm, and the information on migration, I view this as further evidence toward the idea that their may be a generational split for speech rhythm that hovers around those born before the late 1980s/early 1990s and those born after.

In terms of the three speakers involved here, Max, Mezdar and Shorty have mean nPVIV values of 48.8, 52.4, and 52.1, respectively. Solomon, whom they refer to, has a mean nPVIV of 59.0 (see p. 140).

'Staccato' rhythm may therefore actually be two innovations – one led by group second-language acquisition in the mid 1980s, and one led by group second-language acquisition in the late 1990s that targeted the superstrate that the first cohort had established and then began to innovate a new degree of 'staccato' in adolescence. Alternatively or simultaneously, this '*staccato 2.0*' may be what Labov (1972a) has referred to as a *hypercorrection from below* following an incipient *change from below*:

> The changes began as generalizations of the linguistic form to all members of the subgroup; we may refer to this stage as *change from below*, that is, below the level of social awareness. The variable shows no pattern of stylistic variation in the speech of those who use it, affecting all items in a given word class. The linguistic variable is an *indicator*, defined as a function of group membership.
>
> Succeeding generations of speakers within the same subgroup, responding to the same social pressures, carried the linguistic variable further along the process of change, beyond the model set by their parents. We may refer to this stage as *hypercorrection from below*. The variable is now defined as a function of group membership and age level. (Labov 1972a:178)

It is probable that a coalescence of the events discussed above – spikes in migration and segregation – motivated the inception of what I would glibly refer to as 'Rinkeby Swedish 2.0'. When one considers the findings in Chapter 6 that one's childhood school diversity correlates strongly to nPVIV (Model 9, p. 173), the case strengthens for this interpretation. This would not be unlike the generational split that exists in London between the 1980s/1990s London Jamaican (Hewitt 1986) and the 2000s Multicultural London English (Cheshire et al. 2011). Emblematic features of the latter, such high-front FACE, high-back GOAT, and fronted GOOSE, can be understood as hypercorrections from below of the former (Cheshire et al. 2011:160–162)[5]. Further, Rampton (2006) has made an argument that the neoliberal forces of late modernity, including the British school reform of 1988, contributed to the inception of similar multiethnic speech styles in the Midlands of England (2006:3–38).

It is difficult to separate the pressure that new enclosures (like segregation and class stratification) put on language from the influences of fresh contact-driven inputs. This is especially the case when the two are contemporaneous. But group second-language acquisition is not mutually exclusive to the pressures of nonconformity[6]. Just as Kerswill and Williams (1994) have conceptualized the selection of features during koinéization as a change from below, Sharma and Sankaran (2011) have argued that the shedding of /ʈ/ among first-generation British Asians reflects its evolution from indicator to marker (2011:417). The data here implies the same for a contact feature like low-alternation prosody. In response to social pressures like segregated schools that would have siloed off mainstream inputs – combined with a wave of *new* inputs from new pupils caused by the second immigration spike – staccato 1.0 was ratified and escalated by adolescents in the late 1990s and early 2000s to staccato 2.0.

However, 'not all staccatos are created equally'. In the next section, it becomes evident that staccato 2.0 is not merely a more extreme rendition of staccato 1.0. Where staccato 1.0 relied on the reduction of accented vowels, staccato 2.0 appears to rely more on the enlargement of unstressed vowels.

4 Apparent change on the phonology of rhythm: Different cohorts, different strategies

How staccato rhythm is achieved is heavily influenced by age. Older *invandrare* speakers manipulate accented long-vowel combinations. Younger speakers also manipulate accented long-vowel combinations, but to a lesser degree. They also innovatively enlarge unstressed vowels along the dimensions of duration, intensity, and f_0. In this section, I conducted a conditional-tree analysis (*ctree*) to identify which groups are operating what changes on the

[5] I would also add that just as London Jamaican contains many noticeable Cockney features, *Rinkeby Swedish 1.0* contains Södersnack features, the most noticeable of which is the *Stockholm's e* (refer to p. 47 for a description of this feature).

[6] See Labov (2001:513) on the *Nonconformity Hypothesis*.

phonology. From the model output, I identify four key vowel-pair combinations that constitute most of the variation. At the end of the analysis, I model seven social dialects according to the ctree output.

Conditional inference tree analyses constitute the basis for the Random Forest analysis I conducted in Chapter 8. A conditional inference tree groups the response variable into clusters and creates a hierarchical prediction path that leads to each response cluster. They offer important information about change-points or what Tagliamonte and Baayen (2012:159) refer to as 'binary splits'.

> A conditional inference tree provides estimates of the likelihood of the value of the response variable [...] based on a series of binary questions about the values of predictor variables. [...] The algorithm works through all predictors, splitting the data into subsets where justified, and then recursively considers each of the subsets, until further splitting is not justified. In this way, the algorithm partitions the input space into subsets that are increasingly homogeneous with respect to the levels of the response variable. (Tagliamonte & Baayen 2012:159)

Although conditional inference trees are typically deployed for categorical response variables, continuous response variables are also precedented. For example, Grömping (2009) demonstrated this in his investigation of fertility rates in Swiss provinces. His conditional inference-tree analysis broke up fertility rates into four meaningful clusters and demonstrated that provincial demographics combined at unique thresholds to predict one of four possible fertility clusters (2009:310).

Using the *partykit* package in *R*, I ran a Random Forest *ctree* analysis with nPVIV in CASUAL as the response variable. The predictors were the internal factors listed on page 218 and birth year[7], racialization, and social class with the following call:

> nPVIV ~ ACCENT_1_L + ACCENT_1_S + ACCENT_2_L + ACCENT_2_S + UNSTRESS_L + UNSTRESS_S + CODA_R + PHRASE_FIN + LEX_FREQ + SPCH_RATE + RACE + CLASS + BIRTH_YEAR

Figure 9.9 contains the output. Decision nodes are numbered 1 through 103. Most of the nodes reaffirm what has been discussed in Chapter 6. In terms of age, however, they offer some interesting additional insights. Decision nodes 7, 31, 34, 37, 60, 95, and 97 – all highlighted in pink – concern birth year and reveal information about apparent change within the phonology of rhythm. Social class nodes are highlighted in blue. Racialization nodes are highlighted in yellow. The tree clusters the variation into nine main types of vowel pairs. They are listed below alongside the number of vowel pairs (and percentage) they cover. The relevant section that deals with them is also provided. This analysis will only cover the four most-frequently occurring vowel-pair combinations, listed at the top of the following list.

[7] Note that birth year is used instead of age in 2017, because the predictors do not need to be normalized to the same 100-point scale.

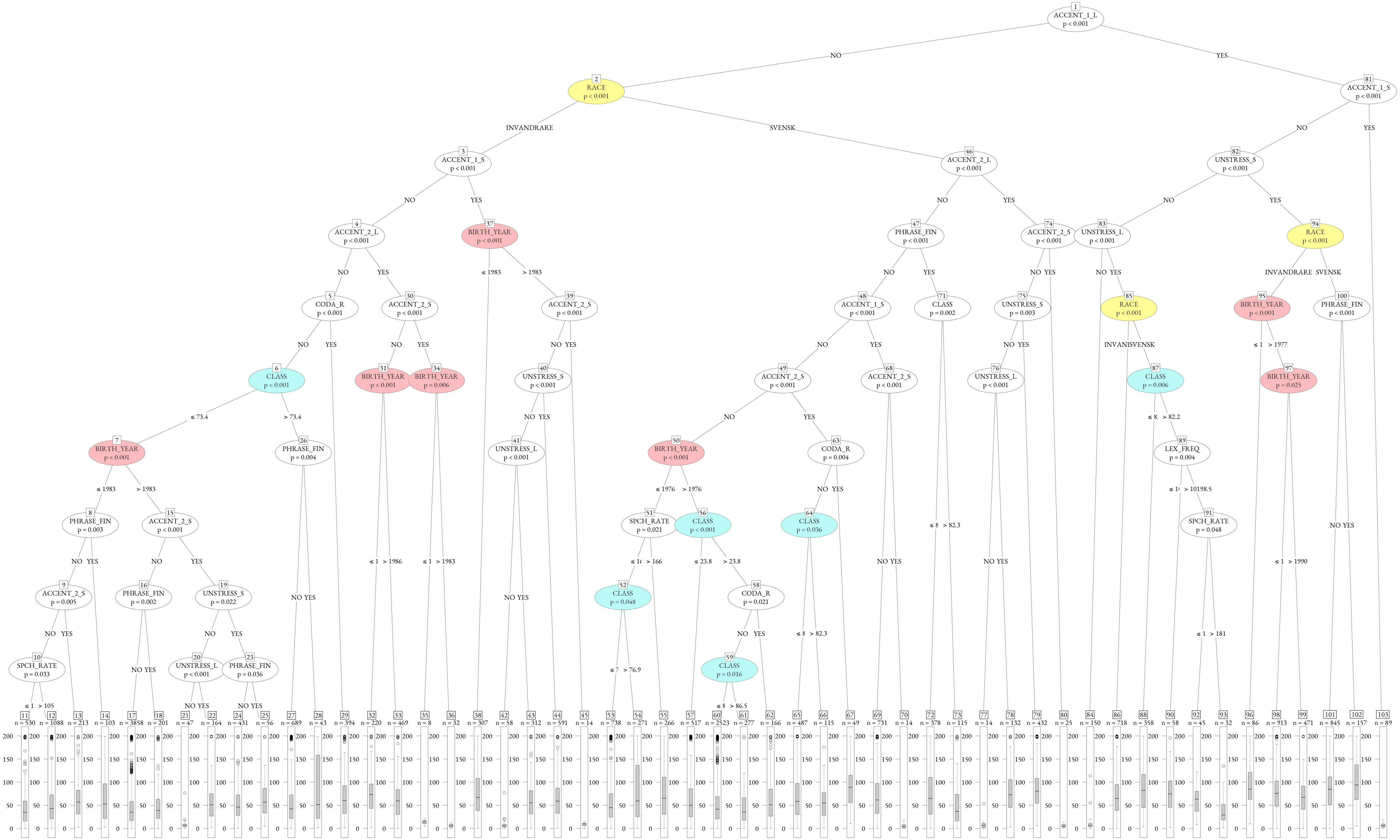


Figure 9.9: Conditional inference tree with nPVIV in CASUAL speech as the response variable. Social predictors are colored: age - pink, class - blue, racialization - yellow

	vowel-pair combinations in dataset	count	pct
Section 4.1	unstressed vowels next to unstressed vowels (phonological quantity irrelevant)	n = 11 722	55.8%
Section 4.2	Accent-1 long vowels next to unstressed short vowels	n = 2 472	11.8%
Section 4.3	Accent-1 short vowels next to unstressed short vowels	n = 1 326	6.3%
Section 4.4	Accent-2 short vowels next to unstressed short vowels	n = 1 286	6.1%
	Accent-1 long vowels next to unstressed long vowels	n = 1 211	5.8%
	Accent-2 long vowels next to unstressed short vowels	n = 972	4.6%
	Accent-1 short vowels next to unstressed long vowels	n = 687	3.3%
	Accent-2 short vowels next to unstressed long vowels	n = 485	2.3%
	Accent-2 long vowels next to unstressed long vowels	n = 268	1.3%

Under the heading for each subsection are charts that contain median normalized duration, intensity and f_0 for the social categories that the conditional inference tree determined were relevant. Note that medians are used to safeguard against skewing (Thomas & Carter 2006:343).

There is one discrepancy between the visual description in Figure 9.9 and the numerical values handled in the analysis. The *partykit* graphics package erroneously centers the black line in the boxplots at the mean nPVIV value rather than the median. This numerical report provided by the analysis, however, uses median nPVIV values. This is what I also report here.

4.1 Unstressed vowels next to unstressed vowels

e.g., *ˈuppsatta plaˈkater* [a pla]

	invandrare higher class	*svensk* born 1976 & younger higher class	middle class	lower class	*invandrare* born 1983 or older	*invandrare* born after 1983
median normalized duration (ms)	54.5	53.1	55.0	58.1	52.5	57.8
median normalized intensity (dB)	53.8	53.8	53.5	54.4	53.6	53.8
median normalized f_0 (Hz)	123.0	123.8	120.3	114.6	118.1	121.7
EFI	361	354	354	362	332	378

Table 9.2: UNSTRESSED VOWELS that occur next to unstressed vowels

The ctree did not find any significant difference between unstressed phonologically long and unstressed phonologically short vowels when they are adjacent to each other. These constitute the single largest vowel-pair category – 55.8% of the corpus.

This analysis concerns nodes 11, 12, 17, 27, 53, 54, 55, 57, 60, and 61. The model clusters the speakers in a manner that is hard to read. Therefore, I have consolidated these clusters in Table 9.3. Several trends are apparent here. The trend for older *svensk* speakers is that higher class corresponds with higher alternation, and lower class corresponds with lower alternation. For younger *svensk* speakers, this trend appears to be reversed. Younger *invandrare* have more staccato speech than older (consistent with all the earlier findings). Finally, and

	UNSTRESSED VOWELS that occur next to unstressed vowels									
							middle & lower class			
	svensk born before 1976			*svensk* born 1976 & younger			*invandrare*	*invandrare*		
	faster speech		slower				born	born 1983 & older		higher
	higher class	middle & lower class	speech all classes	higher class	middle class	lower class	after 1983	faster speech	slower speech	class *invandrare*
median nPVIV	83.3	59.0	83.0	43.5	52.2	66.5	42.2	43.4	53.3	45.2
observations *n*	271	738	266	277	2 523	517	3 858	530	1 088	689
node in Fig 9.9	54	53	55	61	60	57	17	11	12	27

Table 9.3: Breakdown of the predicted median nPVIV for unstressed vowels for the various clusters of speakers established by the conditional inference tree in Figure 9.9 (p. 248).

most interesting for this analysis, the trends for younger *invandrare* are such that they have similar nPVIV values to younger higher-class *svensk* speakers. Given how differently these groups sound, and how prevalent this vowel combination is, how can this be the case? The breakdown in Table 9.2 of individual vowel properties helps explain how.

Whereas higher-class *svensk* speakers may produce the same intervocalic contrast as *invandrare* speakers born after 1983, the constituent vowels are very different in terms of size (EFI=354 and 378, respectively). Young lower-class *invandrare* speakers have unstressed vowels that are 6.8% bigger in EFI and 8.9% longer in duration than higher-class *svensk* speakers. Older *invandrare* speakers fall in-between.

Higher-class *invandrare* speakers show a very similar median nPVIV to higher-class *svensk* speakers in Table 9.3 (45.2 to 43.5, respectively). The two groups also have nearly identical normalized vowel measurements. This is in line with the findings in Chapters 6 and 8 that show racialization to have little predictive power on speech rhythm in the higher classes. It also is consistent with the literature reviewed on 'elites' in Section 4 (pp. 31–35) that has implied that they had become an embedded part of Stockholm's elite.

4.2 Accent-1 long vowels next to unstressed short vowels

e.g., *pla'kat* [pla ¹kɑːt]
'dagen [¹dɑː gɛn]

	invandrare born 1977 or older ('Kevin')	ALL *svensk*	*invandrare* born 1977–1990	*invandrare* born after 1990
median normalized duration (ms)	113.4	110.6	104.6	108.1
median normalized intensity (dB)	57.3	55.7	56.9	56.2
median normalized f_0 (Hz)	136.9	135.1	133.5	137.2
EFI	890	832	795	834

Table 9.4: ACCENT-1 LONG VOWELS that occur next to unstressed short vowels

When an accent-1 long vowel falls adjacent to an unstressed short vowel, the model predicts a divide between all *svensk* and *invandrare* speakers. There are three age groupings among *invandrare* speakers: those born before 1977, those born 1977–1990, and those born after 1990.

	invandrare born 1977 or older ('Kevin')	ALL *svensk*	*invandrare* born 1977–1990	*invandrare* born after 1990
median normalized duration (ms)	52.8	53.5	53.9	59.6
median normalized intensity (dB)	52.4	53.4	53.6	54.1
median normalized f_0 (Hz)	110.1	121.3	120.4	124.1
EFI	305	347	348	400

Table 9.5: UNSTRESSED SHORT VOWELS that occur next to accent-1 long vowels

Invandrare born before 1977 consists of just one speaker, 'Kevin' and concerns node 96. His nPVIV is actually higher than the median nPVIV for *svensk* speakers at 92.7 (node 96). *Svensk* speakers concern node 101 and have a median nPVIV of 84.7. *Invandrare* born between 1977 and 1990 concern node 98 and have a median nPVIV of 76.8 – lower than both Kevin and the *svensk* speakers. *Invandrare* born after 1990 concern node 99 and have an even lower median nPVIV of 69.4.

For the EFI of accent-1 long vowels, shown in Table 9.4, Kevin (older *invandrare*) has the largest vowels (890), followed by younger *invandrare* men (834), followed by *svensk* men (832), followed by the middle cohort of *invandrare* men (795). For the EFI of adjacent unstressed short vowels, shown in Table 9.5, Kevin (older *invandrare*) has the smallest vowels (305), followed by *svensk* men (347), followed by the middle cohort of *invandrare* men (348), followed by younger *invandrare* men with the largest unstressed long vowels (400).

The key finding within this combination is that the youngest *invandrare* group achieves staccato rhythm by means of a different strategy than the middle *invandrare* group. The middle group reduced the accented vowel to achieve its staccato effect. The younger group partially restored the EFI of the accented vowel and then enlarged the adjacent unstressed vowel. Note how in the previous section, this younger group also enlarges unstressed vowels next to unstressed vowels by a considerable amount.

This constitutes a major shift in how the staccato effect is achieved. Rather than continuing the reduction of the accented vowel, the accented vowel is partly restored, and the younger group hyper-corrects the staccato effect by making the adjacent unstressed vowel larger. This is primarily achieved by increasing the vowel's duration by 10.6% (59.6 versus 53.9).

4.3 Accent-1 short vowels next to unstressed short vowels

e.g., *ˈbilden* [^{1}bɪlː dɛn]
om ˈbilden [ɔm ^{1}bɪlː]

	ALL *svensk*	*invandrare* born 1983 or older	*invandrare* born after 1983
median normalized duration (ms)	83.0	85.3	90.7
median normalized intensity (dB)	55.9	55.5	55.6
median normalized f_0 (Hz)	136.6	147.1	134.9
EFI (Hz)	633	697	680

Table 9.6: ACCENT-1 SHORT VOWELS that occur next to unstressed short vowels

	ALL *svensk*	*invandrare* born 1983 or older	*invandrare* born after 1983
median normalized duration (ms)	51.6	48.8	56.8
median normalized intensity (dB)	54.0	52.9	53.8
median normalized f_0 (Hz)	123.6	117.7	122.1
EFI	344	304	373

Table 9.7: UNSTRESSED SHORT VOWELS that occur next to accent-1 short vowels

When an accent-1 short vowel falls adjacent to an unstressed short vowel, the model predicts a divide between all *svensk* and *invandrare* speakers. There are two age groupings among *invandrare* speakers: those born before 1983 and those born after 1983. The dynamic resembles that described in the previous section. Younger *invandrare* speakers manipulate the unstressed vowel, and older *invandrare* speakers manipulate the accented vowel.

Svensk speakers concern node 69 and have a median nPVIV of 69.8 (node 69). *Invandrare* born in 1983 or before concern node 38 and have a median nPVIV of 75.5. *Invandrare* born after 1983 or concern node 44 and have a median nPVIV of 62.8.

For the EFI of accent-1 short vowels, shown in Table 9.6, *svensk* men have the smallest vowels (633), followed by younger *invandrare* men (680), followed by older *invandrare* men (697). For the EFI of the adjacent unstressed short vowels, shown in Table 9.7, younger *invandrare* men have the largest vowels (373), followed by *svensk* men (344), followed by older *invandrare* men (304).

What appears to have happened here resembles the transformation that occurred with accent-1 long vowels except in the other direction. Older *invandrare* began to enlarge the accent-1 short vowels, and younger *invandrare* have partially restored their size and have, in turn, enlarged the adjacent unstressed short vowel by a great deal. This has resulted in a lower nPVIV (62.8, node 44, compared to 75.5, node 38).

Recall in Chapter 7 that I concluded that the staccato effect was more a function of the shortening of prominent long vowels than the lengthening of prominent short vowels (pp. 182–189). The breakdown here helps explain why. *Svensk* speakers and younger *invandrare* speakers end up having more similar nPVIV values (69.8 and 62.8, respectively) than older *invandrare* speakers (75.5).

4.4 Accent-2 short vowels next to unstressed short vowels

e.g., *skriver ˈuppsats* [vɛ ^{2}rɵpː]
ˈuppsatt [^{2}ɵpː sat]

When an accent-2 short vowel falls adjacent to an unstressed short vowel, the model predicts a class divide between the higher-class *svensk* and the remaining *svensk* speakers. It also predicts a class divide between the higher-class *invandrare* and the remaining *invandrare* speakers. Within the remaining *invandrare* speakers – lower- and middle-class – it also predicts an age divide between those born in 1983 and earlier and those born after 1983.

	higher-class *svensk*	lower & middle-class *svensk*	higher-class *invandrare*	lower & middle-class *invandrare* older ≤1983	younger >1983
median normalized duration (ms)	79.0	79.0	86.4	78.9	85.5
median normalized intensity (dB)	53.9	55.4	55.3	54.8	55.0
median normalized f_0 (Hz)	128.5	144.1	124.8	133.8	131.4
EFI	548	630	597	578	619

Table 9.8: ACCENT-2 SHORT VOWELS that occur next to unstressed short vowels

	higher-class *svensk*	lower & middle-class *svensk*	higher-class *invandrare*	lower & middle-class *invandrare* older ≤1983	younger >1983
median normalized duration (ms)	61.9	67.7	60.1	56.0	65.0
median normalized intensity (dB)	53.2	54.3	53.4	53.3	53.8
median normalized f_0 (Hz)	120.6	122.3	115.8	116.5	120.9
EFI	398	450	372	348	423

Table 9.9: UNSTRESSED SHORT VOWELS that occur next to accent-2 short vowels

Higher-class *svensk* men concern node 66. They are predicted to have a mean nPVIV of 55.0 (node 66). Lower- and middle-class *svensk* men concern node 65. They are predicted to have a mean nPVIV of 70.0. Higher-class *invandrare* speakers as defined here include only Majeed (born 1986) and Parviz (born 1992) and concern node 27. They are predicted to have a mean nPVIV of 55.2. Older lower- and middle-class *invandrare* speakers concern node 13. They are predicted to have a mean nPVIV of 60.8. Younger lower- and middle-class *invandrare* speakers concern node 24. They are predicted to have a mean nPVIV of 52.6.

For the EFI of accent-2 short vowels, shown in Table 9.8, higher-class *svensk* men have the smallest vowels (548), followed by older lower- and middle-class *invandrare* men (578), followed by higher-class *invandrare* men (597), followed by younger lower- and middle-class *invandrare* men (619), followed by lower and middle-class *svensk* men (630). For the EFI of adjacent unstressed short vowels, shown in Table 9.9, lower- and middle-class *svensk* men have the largest vowels (450), followed by younger lower- and middle-class *invandrare* men (423), followed by higher-class *svensk* men (398), followed by higher-class *invandrare* men (371), followed by older lower- and middle-class *invandrare* men (348).

These patterns are the hardest to understand and could reflect a shortcoming in the methodology. In this analysis I decided to collapse pretonic and postonic vowels into the same category. This was done to simplify an already quite complex analysis. However, recall my review of accent 2 on pages 45–47. Accent-2 vowels have a delayed peak that sometimes carries onto the post-tonic vowel, especially if they are BIG. In that sense, future analyses ought to handle postonic and pretonic vowels – at least surrounding accent 2 – as separate phenomena.

Accent-2 vowels constitute only a small percentage of the total vowel inventory (14.2 %), so my position is that the patterns outlined in the prior sections are still valid.

4.5 Seven social dialects: A rhythm summary

For some phonological-category combinations, the ctree analysis identified critical breaks and clusters between certain age groups, social-class groups, and/or between *invandrare* and *svensk* identities. For other phonological-category combinations, those same splits showed either no social differentiation in production or an entirely different social stratification. When the various phonological categories are added together, rhythm shows a seven-way split, which corresponds to seven distinct social groups.

Simulations of these rhythmic patterns are constructed in Figures 9.10 and 9.11 for the phrase *tidigare på dagen satt jag uttråkad* (earlier in the day I was sitting around bored). The dialects are (1) older[8] higher-class, (2) younger higher-class, (3) older lower and middle-class *svensk*, (4) younger middle-class *svensk*, (5) younger lower-class *svensk*, (6) older lower and middle-class *invandrare*, (7) and younger lower and middle-class *invandrare*. Figure 9.10 offers a visual reconstruction of duration and f_0 to show how the components of prominence have evolved in apparent time. Figure 9.11 offers a similar reconstruction for duration and intensity for the same phrase.

Explaining the establishment of Figures 9.10 and 9.11

The aim of Figures 9.10 and 9.11 is to offer a visual representation of rhythm that cuts across social class, racialization and age according to the most significant nPVIV splits found in the ctree analysis.

My establishment of seven rhythmic patterns is taken directly from the significant distinctions identified in the ctree analysis in Figure 9.9. Let us, for example, begin with the first syllable *ti-* in *tidigare*, which is both lexically stressed in this scenario and contains an accent-2 long vowel. For *svensk* and higher-class speakers, the ctree model predicts no difference for any nPVIV pair that includes this type of vowel. Therefore, I have taken the mean duration of this vowel (112.2 ms) for the the two higher-class subgroups and the three lower and middle-class *svensk* subgroups. These constitute the first five groups in Figures 9.10 and 9.11. The duration of 112.2 ms is represented along the *x*-axis, and the mean f_0 of 130.4 Hz is represented along the *y*-axis in Figure 9.10. The mean intensity of 55.5 dB is represented in the *y*-axis in Figure 9.11. For the last two *invandrare* groups in the figures, however, the ctree model predicts a significant difference between the two. For older middle and lower-class *invandrare*, the mean duration is 109.8 ms with a mean f_0 of 122.8 Hz and a mean intensity of 55.6 dB. For younger middle and lower-class *invandrare*, the mean duration is 105.6 ms with a mean f_0 of 128.3 Hz and a mean intensity of 56.3 dB.

Turning to the next syllable *di-*, this is an unstressed long vowel that sits adjacent to an accent-2 long vowel. The ctree model predicts a similar division: no difference within the first five groups but a difference for the last two *invandrare* groups.

[8] Recall that 'older' and 'younger' are with respect to the speakers in this dataset and correspond with those born in the late seventies/early eighties versus those born after.

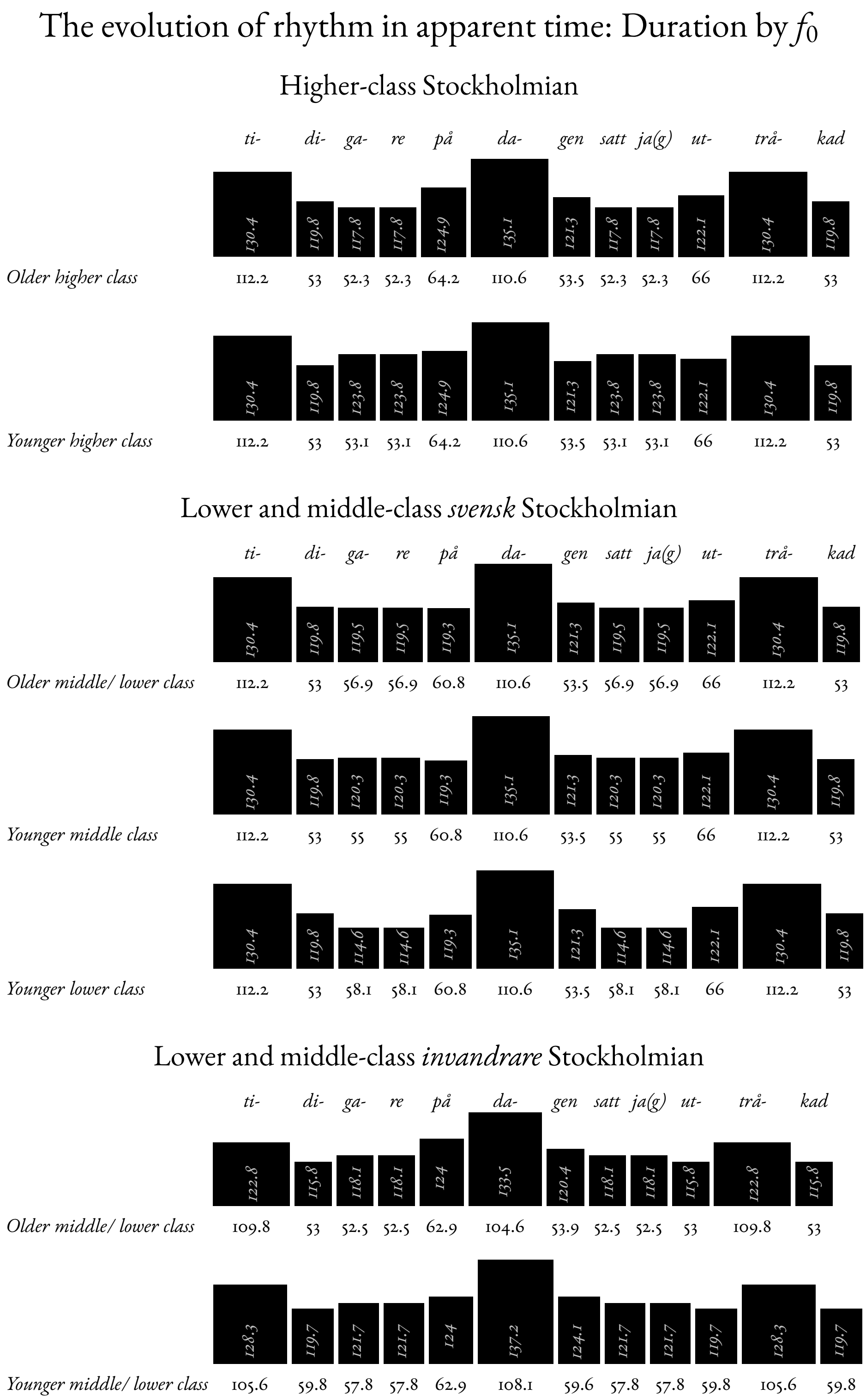


Figure 9.10: Evolution of rhythm in apparent time for seven Stockholm sociolects (duration and fundamental frequency), reconstructed from the Random Forest model in Figure 9.9, for the phrase '*tidigare på dagen satt jag uttråkad*' ('*earlier in the day I was sitting around bored*'). Duration is the horizontal parameter, and f_0 is the vertical parameter. Duration is exactly to scale; f_0 has been magnified fourfold with the bottom 100 Hz truncated.

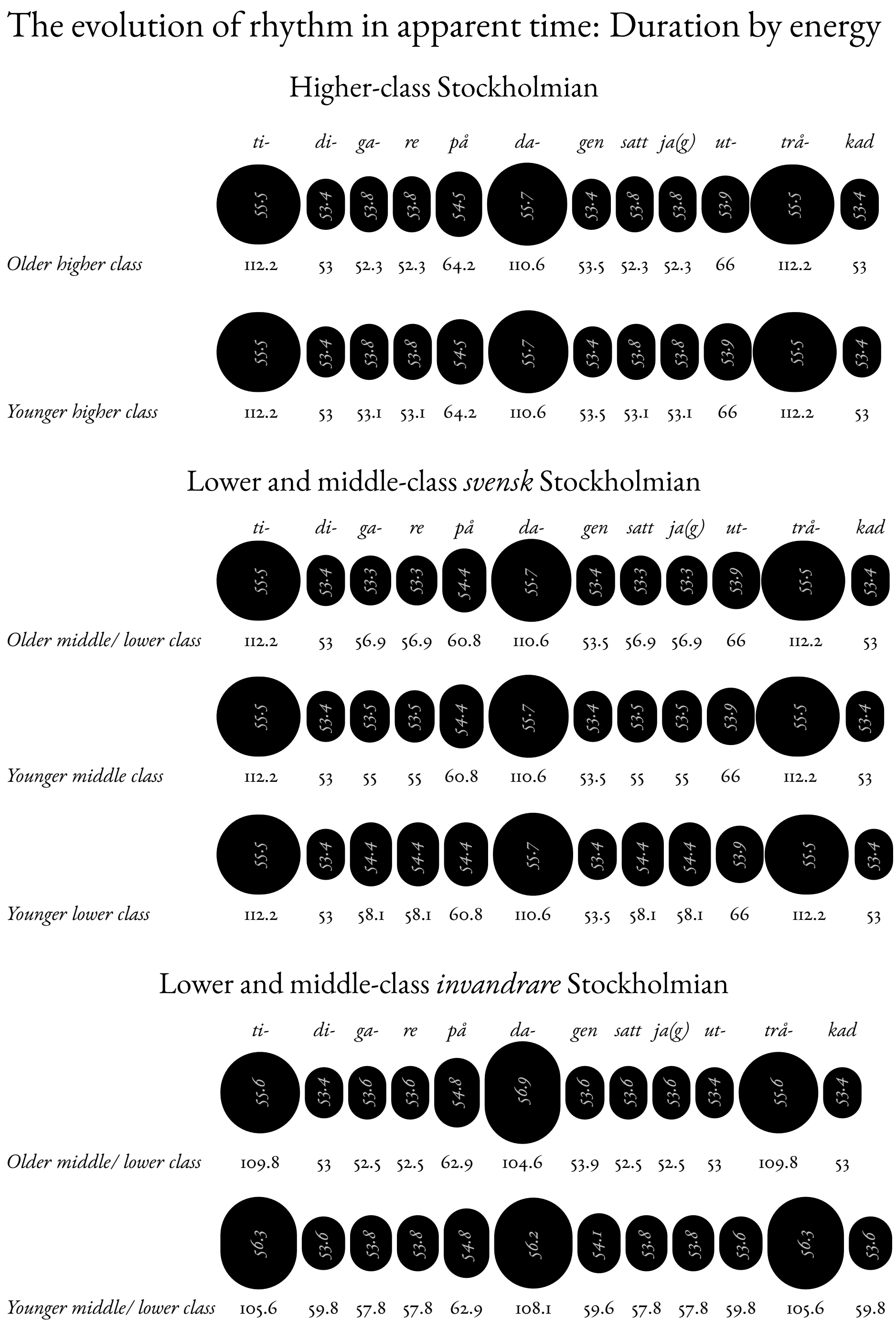


Figure 9.11: Evolution of rhythm in apparent time for seven Stockholm sociolects (DURATION and INTENSITY), reconstructed from the Random Forest model in Figure 9.9, for the phrase 'tidigare på dagen satt jag uttråkad' ('earlier in the day I was sitting around bored'). Duration is the horizontal parameter, and intensity is the vertical parameter. Duration is exactly to scale; intensity has been truncated by 45 dB and then squared to approximate the visualization algorithm in Praat (intensity is squared and then convolved with a Gaussian analysis window).

If we turn to the third syllable *ga-*, which is an unstressed short vowel that sits adjacent to an unstressed vowel, the model predicts an entirely different social distribution than for the preceding two vowels. It is here where the seven-way split is the most meaningful. Older higher-class and younger higher-class show a significant difference in nPVIV values for this specific combination, so I separated out the means for duration, f_0, and intensity for these two groups. Older higher-class speakers have a mean duration of 52.3 ms, a mean f_0 of 117.8 Hz, and a mean intensity of 53.8 dB. Younger higher-class speakers have a mean duration of 53.1 ms, a mean f_0 of 123.8 Hz, and a mean intensity of 53.8 dB. Older middle and lower-class *svensk* speakers have a mean duration of 56.9 ms, a mean f_0 of 119.5 Hz, and a mean intensity of 53.3 dB. Younger middle-class *svensk* speakers have a mean duration of 55 ms, a mean f_0 of 120.3 Hz, and a mean intensity of 53.5 dB. Younger lower-class *svensk* speakers have a mean duration of 58.1 ms, a mean f_0 of 114.6 Hz, and a mean intensity of 54.4 dB. Moving to the final two groups, older middle and lower-class *invandrare* have a mean duration of 52.5 ms, a mean f_0 of 118.1 Hz, and a mean intensity of 53.6 dB. Younger middle and lower-class *invandrare* have a mean duration of 57.8 ms, a mean f_0 of 121.7 Hz, and a mean intensity of 53.8 dB.

For the remaining syllables, the distributions continue to vary. The syllable *på* splits three ways; *da-* splits four ways; *gen* splits three ways; both *satt* and *ja(g)* split seven ways; both *ut-* and *trå* split three ways; and, finally, *kad* splits three ways[9].

Key visual interpretations from Figures 9.10 and 9.11

The reduction of stressed vowels in older lower and middle-class *invandrare* speech is clearly seen, as is their partial restoration in younger lower and middle-class *invandrare* speech. The enlargement of unstressed vowels in younger lower and middle-class *invandrare* speech is also quite apparent. Notably, the durational reduction and the frequency reduction on accented vowels seem to be accompanied by an escalation in intensity. Both *invandrare* varieties have overall higher normalized intensity on their accented vowels than the other five varieties. My interpretation is that the first versions of Rinkeby Swedish ('Rinkeby Swedish 1.0') involved a reduction of prominence – primarily via duration – for accented vowels, which rendered a mildly staccato effect. Late-modern Swedish multiethnolect ('Rinkeby Swedish 2.0'), on the other hand, has relied more on an increase in prominence for unstressed vowels – the most frequent type of vowel (55.8%) – to produce the heavier staccato effect.

Younger higher-class speech has an upwards climb toward stress in f_0 and intensity, which may be why its durationally short unstressed vowels do not perceptually come across as staccato (at least to my ears). In contrast, younger lower-class *svensk* speech shows a 'troughing' of f_0 between stresses that is accompanied by a decrease in intensity after the accent followed by an increase in intensity simultaneous to the f_0 trough. The f_0 pattern of the younger lower-class *svensk* resembles that of the older higher-class speakers, but the duration and in-

[9]Although the older middle and lower-class *invandrare* have the same mean duration of 53 ms and the same mean intensity of 53.4 dB for *kad* as the preceding five social groups, f_0 differs: 119.8 Hz for the preceding five social groups and 115.8 for older middle and lower-class *invandrare*.

pseud.	profession	PC1 social class	ctree grouping	median nPVIV of UNSTRESSED next to UNSTRESSED vowels	median nPVIV of ACCENT-2 SHORT next to UNSTRESSED vowels	ctree grouping	median nPVIV of ACCENT-1 LONG next to UNSTRESSED SHORT vowels	median nPVIV of ACCENT-1 SHORT next to UNSTRESSED SHORT vowels	median nPVIV of all four pair types
Reman	social worker	0.0		45.1	45.7		60.9	48.8	46.2
Malik	cook at cafeteria, street food	1.1		41.9	37.3		49.0	49.2	42.3
Max	youth worker	4.4		40.0	30.9		41.4	47.8	40.4
Antonio	caretaker, ex drug dealer	7.6		36.2	37.6		40.4	28.7	36.4
Hayder	caretaker, bouncer	19.0		44.2	50.7		43.4	50.6	44.6
Dawit	group home assistant	19.8		42.6	44.9		46.1	47.1	43.0
Solomon	youth worker	20.1		45.5	50.0		53.0	52.8	46.6
Murad	group manager, furniture retail	21.7		36.4	35.4		39.8	39.0	37.7
Abel	photographer, personal trainer	25.4	low	44.0	48.0		53.6	41.8	45.0
Mateo	waiter, bartender	31.6		45.5	40.6	all together	54.4	71.8	47.4
Reza	waiter, napropathy student	34.8		43.3	47.3		53.1	46.0	45.4
Mezdar	youth worker (certified)	37.5		43.9	36.1		46.3	46.9	43.5
Jocke	athletic coach, personal trainer	42.2		42.7	44.8		55.7	42.3	43.6
Sohrab	web architect	48.6		50.4	43.5		61.7	75.0	52.4
Shorty	social worker	51.2		42.8	45.6		67.2	41.9	46.1
Kevin	organisational consultant	64.0		47.6	54.5		61.2	51.1	49.4
Tarik	field assistant; athletic coach	73.4		42.5	37.2		51.8	42.5	42.8
Majeed	doctor	76.2	high	49.8	55.6		56.7	59.4	51.9
Parviz	business owner, physician	91.1		46.6	41.9		62.9	57.1	48.5

Table 9.10: Median nPVIV values per speaker of the four largest vocalic-pair types. Speaker pseudonym, profession, and social class are listed along with how the speaker was grouped in the ctree analysis. Total medians are calculated as well in the far right column. 'Anomalous' nPVIV values are shaded in gray. Medians instead of means are reported in accordance with Thomas and Carter (2006:343).

tensity patterns between the two are quite different. The older higher class has more reduction for unstressed vowels in regards to both duration (52.3 ms) and intensity (53.8 dB) than the younger lower-class *svensk* has (58.1 ms and 54.4 dB, respectively). On the other hand, the younger lower-class *svensk* has more f_0 reduction (114.6 Hz) of unstressed vowels than the older higher class does (117.8 Hz).

4.6 Interpreting the weak effects of class among *invandrare*

For some pairs, the ctree show a high social-class split among *invandrare* for [unstressed + unstressed] and [accent-2 short + unstressed short]. Speakers with an index of 73.4 or lower have lower rhythmic alternation for these combinations. These are shown in Table 9.10, and a horizontal line demarcates 73.4. The findings also show no social-class split among *invandrare* for accent-1 long vowels + unstressed short vowels and accent-1 long vowels + unstressed long vowels. These are shown in Table 9.10.

Table 9.10 also shows medians for just these four pair categories. 'Anomalously' high-nPVIV values are shaded in gray. Each of these speakers will be discussed at a qualitative level here. First, however, I wish to address the overarching issue of social class.

By comparing the *invandrare* occupations in Table 9.10, popular notions of who belongs to what social class imply that my class sample skews toward the lower classes. For example, a social worker is ranked 51.2, and a field assistant and athletic coach is ranked at 73.4. Furthermore, the values 0, 50, and 100 represent the beginning, middle, and end of the first principle component of social class within this dataset instead of within society. Given the professions outlined here, it is possible that speakers who have a first principle component below 73.4 correspond more closely with what is generally considered to be the lower-middle class in Stockholm.

Unfortunately, there are few empirical ways to link my ranking back to 'real' class categories given the fact that the categorization of social class in late-modern Sweden has been under-explored. However, in Britain for example, the lifestyle patterns that match popular and academic conceptualizations of middle class do not correspond with the population midpoint of the social class distribution. For example, Savage et al. (2013) found that the group that most resembles the middle-middle class (entitled the 'established middle class') begins its rank at 69 when Britain's population is normalized to 100. The lower-middle class begins its rank at 59 and ends at 68 (entitled the 'technical middle class') (Savage et al. 2013:12). Therefore, it is important to remember that the term 'middle' has little to with central distribution. Rather, it has to do with the fact that its members fall in-between two more easily-defined groups: low-autonomy workers and gentry.

Returning to the five gray-shaded boxes in Table 9.10, some ethnographic information helps explain why these speakers have median nPVIV values that are somewhat anomalous. Reman is categorized as having the lowest social class background primarily due to his lack of tertiary education and his parents' lack of secondary-school education. His profession, however, is considered lower middle-class by traditional notions of the term. His speech, also is the subject of much metalinguistic commentary by his two best friends, Hayder and Murad. They comment that he speaks 'clean Swedish' because he went to majority white Swedish schools (his *classmates foreign background* score, however, does not reflect this and is similar to Hayder's and Murad's; p. 171). Solomon is part of the older second generation, the same cohort that Kotsinas (1988a) studied. His friends have commented on the fact that his Swedish sounds clean (p. 244). Mateo belongs to the same older generation as Solomon; his nPVIV, however, is even higher than Solomon's. This is likely due the fact that nearly all of his contacts are *svensk* and lower-class (p. 170). Sohrab is interesting because, given his profession and income, he should be closer to Majeed and Parviz. His parents' background, however, have contributed to a lower social-class index. And finally, Tarik has a profession that would otherwise characterize him as quite low on the social-class ranking. However, his other social-class affiliations are quit high. These include a high income and an highly-educated father with a high-prestige occupation.

I offer one final (and somewhat controversial) interpretation for the weak role of social class in the ctree findings. Given the prominence of race in Swedish society, the local multiethnolect might be evolving into a racial dialect with similarities to AAVE whereby even higher-class speakers use certain vernacular features in casual speech (e.g. /r/ absence in Wolfram 1969:116). As I reviewed on pages 31–35, there is significant pressure from the lower end of the class hierarchy for non-white elites to remain 'authentic'. If this sort of pressure is strong enough, it might motivate the production of intermediate vernacular variants among non-white elites. This would explain why some of the more salient vowel combinations (Sections 4.2 and 4.3) show no significant class division among *invandrare* men. It could be that accent-1 pairs are sufficiently salient that elite *invandrare* speakers are commandeering them to do class-attenuated and race-accentuated work.

5 The origin of low alternation in higher-class *svensk* speech

The classed and raced affiliations associated with multiethnolects have driven the focus of this dissertation, but a consistent finding in the analysis has been that ALL speech in Stockholm is becoming more staccato with apparent time. Among *svensk* speakers, the younger middle and higher social classes are in the lead rather than the lower class. This can be seen in Figure 9.1. A question I have, therefore, is: what is causing this change? Could it be due to a trickling diffusion of multiethnolectal prosody[10]?

Some headway can be made into investigating this question because I have data on network structure, which can reveal insights about the diffusion of community-wide trends.

5.1 A literature review on network structure and language change

In Chapter 6 Section 4.1 I reviewed the literature that has shown that network *qualities* are robust predictors of linguistic variation and change (pp. 161–165). Network *structure*, however, is just as important.

Labov (1973) identified network position as key in identifying the source of linguistic variation in a population that generally all belonged to the same social class. He investigated the social network of 94 adolescent boys living in a social housing high-rise in Harlem, New York. Despite their shared address and shared social background, variation was palpable. Boys in this tower who were peripheral to the core peer group and who participated in adult-managed activities were colloquially referred as *lames*. In contrast, boys who were central to networks and part of dense networks (the gangs *T-birds* and *Aces*) barely style-shifted at all (1973:90).

In the same study, Labov (1973) examined the social network of 36 gang members in the neighborhood surrounding the aforementioned apartment tower. He found a direct relationship between network centrality/peripherality with vernacular use/non-use (1973:93).

[10] It was suggested in an earlier review that English use could be contributing to this. However, the literature on rhythm shows English rhythm to have similar or higher contrast than Swedish rhythm (English.: Fuchs 2013:81 and Thomas and Carter 2006; Swedish: Young 2018c:50).

Importantly, he concluded that social-network structure offered an avenue toward securing *representativity* in a field that relies on relatively thin stratified samples:

> At first glance, lames appear to be members of the community; they are much more accessible to the outsider than members are; and the limitations of their knowledge are not immediately evident. But the result as we have seen may be an inaccurate or misleading account of the vernacular culture. [...] Subjects selected with the assistance of teachers, psychologists, or parents are even more heavily biased towards the lame population and unfortunately, the number of studies of Black English made in schools far outnumbers the studies done by direct contact with members in a vernacular context. (Labov 1973:110)

Peripheral speakers are the most accessible to outsiders. They are therefore sources of error for linguists who wish to generalize about vernacular use. Labov found high linguistic variation within an otherwise homogeneous group and identified networks as the key correlate. He therefore convened his analysis by summoning linguists to either investigate networks in their fieldwork or, where network information was unavailable, hedge their conclusions about working-class speech, bearing the *lame* factor in mind (1972b:108–114).

Cheshire (1982) discovered a network dynamic similar to Labov in her sociolinguistic study of adolescent speech in Reading, UK. She was able to construct a network-centrality hierarchy based on the number of names each boy gave as a friend and the number of times a boy's name was given as a friend. The most central boys tended to use the highest percentages of certain nonstandard variants, followed by intermediately-positioned boys, followed by peripherally-positioned boys 1982:90–91. Other features, however, did not correlate with network centrality, which led Cheshire to conclude that only some nonstandard features are markers of peer group status. She also points out that the network in Labov's study was more exhaustive due to the closed and exclusive nature of New York gang affiliation, whereas the networks of working-class British adolescents tended to be more diffuse. This meant that her network model only captured one of numerous interactional aggregations (1982:90).

Milroy (1987) interviewed 23 women and 23 men from Belfast in the late 1970s and assigned them *network-density scores* based on five structural criteria: (1) membership of a high-density territorially-based cluster, (2) substantial kinship ties to the neighborhood, (3) working in the same place as at least two others from the speaker's area, (4) working in the same place of work as at least two others from the same sex from the speaker's area, and (5) voluntary association with workmates in leisure activities (1987:141–142).

She found that older speakers with higher network-density scores were more likely to use vernacular features in their casual speech. Women (and younger men) who worked also tended to have lower network-density scores, and this corresponded highly with the use of more supralocal innovations (Milroy 1987:196–197). Milroy was unable to explore these individuals' networks in depth, but she referred to Granovetter's (1983) theory on cultural and informational diffusion as one possible explanation. It was hypothesized that the speakers accessed supralocal innovations via *bridges* made of *weak ties* (Milroy 1987:199–200).

> What makes cultural diffusion possible, then, is the fact that small cohesive groups who are liable to share a culture are not so cohesive that they are entirely closed; rather, ideas may penetrate from other such groups via the connecting medium of weak ties. (1983:215) [...] I argued that while not all weak ties should be local bridges, all such bridges should be weak ties—an argument central to the assertion that weak ties serve crucial functions in linking otherwise unconnected segments of a network. I argued, for example, that the stronger the tie between two people, the greater the extent of overlap in their friendship circles. (Granovetter 1983:217–218)

In other words, intimate and cohesive relationships juggle around and eventually establish norms. These norms are transmitted via *bridges* to other cohesive networks who then do the same internal juggling. By definition, a *bridge* is two individuals from separate networks with *no mutual contacts*. Since friendship is often accompanied by knowing the friends of one's friends, whenever a weak bridge becomes strong, triadic closure ensues shortly thereafter. This, in turn, *merges* networks rather than bridging them. It has been shown that weak ties are particularly important for job networking, for the diffusion of ideas and innovations, and for the successful spread of political movements (see Granovetter 1983:202 for a comprehensive review).

Bortoni-Ricardo (1985) refers to the diffuseness of networks as 'network integration'. The *network integration index* was calculated based on how connected a speaker's three closest (non-kinship) friends were to the larger pool of 118 community members (1985:130–131). This was done using a matrix calculation of both unidirectional and bidirectional connections (1985:167). Speakers whose friends were less connected to the community had higher *network integration* and lower network insulation. In contrast, speakers whose friends were more connected to the community had lower integration and higher insulation.

It is important to point out that Bortoni-Ricardo (1985) refers to *integrated* networks as the opposite of *insulated* and that neither adequately equate network *density*, per se. In the case of rural migrants in Braslandia, the most material contrast is between kinship and former village ties on one hand and the community ties within the new neighborhood on the other. In this sense, Bortoni-Ricardo (1985) differs from Milroy (1987) who was interested in the density differences between a wider range of social classes (lower-working to lower-middle). Later works, on the other hand, have used the term *integrated* as nearly synonymous with *dense*. For example, Dodsworth and Benton's (2017) definition of *integrated* resembles Bortoni-Ricardo's (1985) definition of *insulated* in the sense that speakers possess various degrees of integration within their *own* kinship-based, local community.

In their study of the attrition of White Southern American English vowels in Raleigh, USA, Dodsworth and Benton (2017) collected data on the schools their participants attended in order to approximate the network properties of the participants during childhood. In other words, two participants who attended the same school were modeled as being connected in a sociogram. The authors found that network embeddedness had contrasting effects on White Southern vowels in two different parts of the city. In the established old Raleigh core, high network embeddedness correlated strongly with the *presence* of White Southern vowels. In North Raleigh, where many affluent migrants from the Northeast reside, network embeddedness correlated strongly with the *absence* of White Southern vowels (2017:392).

Eckert (2000) found that network structure offered an explanation for the especially advanced linguistic features of a group referred to as the 'burned-out burnouts' – a fringe peer group of year-11 students at a Belten Highschool in suburban Detroit, USA. This 'burned-out burnout' network had the most advanced participation in the 'Northern Cities' chain shift of White American English. The network was uniquely *closed* in one respect and *open* in another. It was closed in that only two of its ten members had ties to other classmates in the year-11 class. It was open in that its members had numerous ties outside of the school, including students from other cohorts, other schools and older friends from central Detroit (2000:173). The open characteristic of their network allowed advanced features to diffuse; the closed characteristic allowed these features to take on a flamboyant indexicality by constellating in a closed and exclusive manner (2000:211). The centrality of two of its members may have functioned as a bridge to the 'less burned-out' burnouts.

Fagyal et al. (2010) argue that many network studies render contradictory conclusions. On one hand, as is the case for Milroy (1987), highly central actors are the guardians of conservative features and those with multiple weak ties are the bearers of innovations. On the other hand, as is the case for the gang leaders in Labov (1972b) and the burnouts in Eckert (2000), central actors are the arbiters of innovation. To explore how both could simultaneously be true, Fagyal et al. built a fictitious network of 900 nodes and assigned some nodes high prestige and high connectivity and others low prestige and low connectivity. They simulated the spread of competing variables and found that centrally-located nodes (i.e., those with high prestige and high connectivity) are better conceptualized as *actuator switches* 2010:2073 and are therefore neutral to whether they are propagating (*switch on*) or resisting (*switch off*) change. They refer to Eckert's (2000) Belten Highschool and Milroy's (1987) Belfast to explain this process:

> Centrally-connected charismatic leaders can advance a new and vigorous vowel shift, such as the Northern Cities Shift, in their extended adolescent peer groups in Belten High, Michigan, and thus be perceived as agents of on-going change. They can also propagate local forms in their extended working-class communities in Belfast, resisting the intrusion of mainstream Irish influence, and thus coming across as safekeepers of local norms. The stability of established norms, however, is tied to centers' perceived social influence: in closed, tight-knit social networks their influence depends on how many agents remain sensitive to their

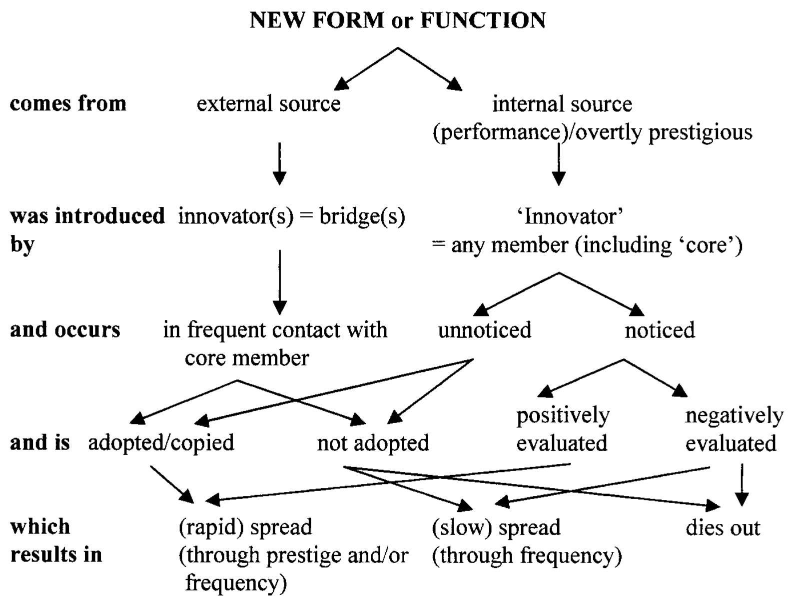


Figure 9.12: Innovation type and change in networks. Reproduced from Bergs (2005:42)

relative prestige over time. (Fagyal et al. 2010:2073)

Bergs (2005) provides a model for this process that I have reproduced in Figure 9.12. Where the form is externally-sourced, it enters via a bridge. Where its source is internal, it emerges via some distinctive move (performance, prestige)[11]. Following its inception and introduction, the form then passes through a center who operates as Fagyal et al.'s (2010) *actuator switch*. An unnoticed form is either readily adopted or not adopted. In the former situation, the well-connected center catapults its rapid spread. In the latter situation, it can spread slowly via an eventual uptake by other members or it can die out. A noticed form that is positively evaluated by a center enjoys rapid spreading as well. An unnoticed form that is negatively evaluated by a center either spreads slowly via an eventual uptake by other members or it dies out.

5.2 Analysis: Network bridges

This analysis was conducted on the 29 speakers who participated in the full social network analysis. Two metrics were calculated for each speaker to reflect two ways of measuring the '*bridges*' that extend from their social networks. 'Bridge' is an arbitrary term, which is why I tested two definitions. The first definition, entitled *Network Bridges3*, refers to the number of contacts in a participant's network who know THREE or fewer other contacts in that same network. The second definition, entitled *Network Bridges4*, refers to the number of contacts

[11]This internally-sourced innovation can be noticed or unnoticed, but it would seem that Bergs makes the assumption that external forms always enter relatively unnoticed. Regardless of whether this assumption is correct or not, the model is helpful for conceptualizing the time sequence of diffusion.

Pseudonym	racialization	class	age	Number of bridges (3 mutual)	Number of bridges (4 mutual)
Reman	*invandrare*	0.0	24	6	6
Malik	*invandrare*	1.1	28	9	9
Max	*invandrare*	4.4	28	-	-
Antonio	*invandrare*	7.6	24	-	-
Hayder	*invandrare*	19.0	25	1	2
Dawit	*invandrare*	19.8	26	-	-
Solomon	*invandrare*	20.1	37	3	3
Murad	*invandrare*	21.7	25	0	0
Abel	*invandrare*	25.4	26	-	-
Mateo	*invandrare*	31.6	36	1	1
Reza	*invandrare*	34.8	25	3	6
Mezdar	*invandrare*	37.5	28	2	2
Jocke	*invandrare*	42.2	28	0	0
Sohrab	*invandrare*	48.6	33	3	4
Shorty	*invandrare*	51.2	26	0	5
Kevin	*invandrare*	64.0	39	1	5
Tarik	*invandrare*	73.4	31	3	3
Majeed	*invandrare*	76.2	30	-	-
Parviz	*invandrare*	91.1	24	16	20

Pseudonym	racialization	class	age	Number of bridges (3 mutual)	Number of bridges (4 mutual)
Per	svensk	14.7	29	9	11
Rasmus	svensk	23.8	34	7	13
Johnny	svensk	26.9	30	2	2
Sven	svensk	34.7	24	-	-
Paul	svensk	38.2	39	5	7
Jesper	svensk	38.7	23	3	4
Richie	svensk	52.8	30	2	8
Thomas	svensk	60.8	32	14	17
Nils	svensk	64.4	39	0	0
August	svensk	73.5	42	-	-
Jonte	svensk	76.9	43	0	4
Martin	svensk	82.2	39	3	3
Gunnar	svensk	82.3	37	3	5
Joseph	svensk	84.1	33	9	9
Jan-Bertel	svensk	86.5	31	14	15
Johan	svensk	88.9	39	5	8
Jan-Axel	svensk	100.0	32	5	5

Table 9.11: Summary of (1) number of 'bridge' contacts, defined by contacts who know three or fewer mutual contacts, and (2) number of 'bridge' contacts, defined by contacts who know four or fewer mutual contacts

in a participant's network who know FOUR or fewer other contacts in that same network. Very few speakers had contacts who only knew TWO or fewer mutual contacts, so this was not used in the analysis. Beyond four contacts, I felt that the ties were not sufficiently bridge-like and were already constituting triadic closure. Therefore, for this dataset, three and four mutual ties constitute a 'sweet spot'.

Refer to Table 9.12 to see an example of each calculation. As discussed in Chapter 6 (p. 169), Hayder provided me with 43 names in his social network interview. To the left side of the table are the qualities of each contact, including gender, occupational status, the language he speaks with them, and so on.

The right side, indicated with 1s and 0s, constitutes the structure of his network. If a contact knows another contact, s/he is marked with a 1. If not, s/he is marked with a 0. Contact 1 is highlighted in gold because that contact only knows three of the contacts whom Hayder also listed. Contact 29 is highlighted in gold because that contact only knows four of the contacts whom Hayder also listed. Therefore, Hayder's metric for *Network Bridges*$_3$ is 1, and his metric for *Network Bridges*$_4$ is 2.

Figure 9.13 shows the social networks of the 29 speakers labeled with their respective numbers of *Network bridges*$_3$. Table 9.11 shows the distribution of all three predictors across the 29-speaker subset.

Hayder

	Name	Network	Sex	Sexuality	Age	Ethnicity	Occupation	SEI	Language	Accent	Closeness	1 Contact_1	2 Contact_2	3 Contact_3	4 Contact_4	5 Contact_5	6 Contact_6	7 250 customers per week	8 Contact_8	9 *Reman*	10 Contact_10	11 Contact_11	12 Contact_12	13 Contact_13	14 Contact_14	15 Contact_15	16 *Murad*	17 Contact_17	18 Contact_18	19 Contact_19	20 Contact_20	21 Contact_21	22 Contact_22	23 Contact_23	24 Contact_24	25 Contact_25	26 Contact_26	27 Contact_27	28 Contact_28	29 Contact_29	30 Contact_30	31 Contact_31	32 Contact_32	33 Contact_33	34 Contact_34	35 Contact_35	36 Contact_36	37 Contact_37	38 Contact_38	39 Contact_39	40 Contact_40	41 Contact_41	42 Contact_42	43 Contact_43	
1	Contact_1	work	male	straight	45	Swedish	Building superintendant	21.8	Swedish	none	not close		1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	← 3 or fewer contacts
2	Contact_2	work	male	straight	23	Bolivian	Building superintendant	21.8	Swedish	Spanish import	close			1	0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Contact_3	work	male	straight	26	Swedish	Building superintendant	21.8	Swedish	none	not close				0	0	0	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Contact_4	work	male	straight	28	Greek	Security Guard	23.8	Swedish	none	not close					1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
5	Contact_5	work	male	straight	28	Swedish	Security Guard	23.8	Swedish	none	not close						1	1	1	1	1	1	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
6	Contact_6	work	male	straight	23	Chilean	Security Guard	23.8	Swedish	none	not close							1	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7	250 customers per week	work	50-50 fe/male	straight	20-30	mostly Swedish	-	55.0	Swedish	none	not close								1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
8	Contact_8	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close									0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
9	*Reman*	gym	male	straight	24	Kurd	Social care / Security guard	24.4	Swedish	Hood	very close										1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	Contact_10	gym	male	straight	35	Swedish	unknown	-	Swedish	none	close											1	1	1	1	1	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	Contact_11	gym	male	straight	28	Swedish	unknown	-	Swedish	none	close												1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	Contact_12	gym	male	straight	28	Swedish	unknown	-	Swedish	none	close													1	1	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	Contact_13	gym	male	straight	23	Kurd	unknown	-	Swedish	Kurdish import	close														1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	Contact_14	gym	male	straight	18	Lebanese	unknown	-	Swedish	Hood	close															1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	Contact_15	gym	male	straight	21	Turk	unknown	-	Swedish	none	close																0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
16	*Murad*	friend	male	straight	25	Egyptian	Manager furniture store	51.6	Swedish	Hood	very close																	1	1	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
17	Contact_17	friend	male	straight	25	Serb	Sales clerk	29.5	Swedish	Hood	very close																		1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	Contact_18	friend	male	straight	25	Tunisian	Fire inspector	57.3	Swedish	Hood	very close																			1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	Contact_19	friend	male	straight	26	Serb	Internet technician	62.5	Swedish	Hood	very close																				1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
20	Contact_20	friend	male	straight	25	Serb	Social care	25.1	Swedish	Hood	close																					1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	Contact_21	friend	male	straight	25	Congolese	Sales clerk	29.5	Swedish	Hood	very close																						1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	Contact_22	friend	male	straight	25	Maroccan	Warehouse worker	16.4	Swedish	Hood	very close																							1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	Contact_23	friend	male	straight	25	Ghanian	Software developer	62.5	Swedish	none	very close																								0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	Contact_24	family	male	straight	58	Kurd	Home care	25.1	Kurdish	Import	very close																									1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
25	Contact_25	family	female	straight	55	Kurd	Assistant nurse	56.0	Kurdish	Import	very close																										1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
26	Contact_26	family	male	straight	23	Kurd	Social care	25.1	Kurdish / Swedish	Hood	very close																											1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
27	Contact_27	family	female	straight	18	Kurd	University student	-	Kurdish / Swedish	none	very close																												1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
28	Contact_28	family	male	straight	90	Kurd	retired	-	Kurdish	Import	very close																													0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
29	Contact_29	gym	female	straight	30	Kurd	Attorney	-	Swedish	none	not close																														0	0	0	0	0	0	0	0	0	0	0	0	0	0	← 4 or fewer contacts
30	Contact_30	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close																															1	1	1	1	1	1	1	1	1	1	1	1	1	
31	Contact_31	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close																																1	1	1	1	1	1	1	1	1	1	1	1	
32	Contact_32	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close																																	1	1	1	1	1	1	1	1	1	1	1	
33	Contact_33	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close																																		1	1	1	1	1	1	1	1	1	1	
34	Contact_34	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close																																			1	1	1	1	1	1	1	1	1	
35	Contact_35	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close																																				1	1	1	1	1	1	1	1	
36	Contact_36	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close																																					1	1	1	1	1	1	1	
37	Contact_37	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close																																						1	1	1	1	1	1	
38	Contact_38	work	male	straight	25	Invandrare	Security Guard	23.8	Swedish	Hood	not close																																							1	1	1	1	1	
39	Contact_39	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close																																								1	1	1	1	
40	Contact_40	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close																																									1	1	1	
41	Contact_41	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close																																										1	1	
42	Contact_42	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close																																											1	
43	Contact_43	work	male	straight	25	Swedish	Security Guard	23.8	Swedish	none	not close																																												

Network bridges$_3$ - 3 or fewer mutual contacts = 1
Network bridges$_4$ - 4 or fewer mutual contacts = 2

Table 9.12: Sample calculation of 'bridge contacts' for participant Hayder.

I built two mixed-effects regression models that each are adaptations of the base call used in Model 1 from Figure 8.3 (p. 221):

> nPVIV ~ ACCENT_1_LONG + ACCENT_1_SHORT + ACCENT_2_LONG + ACCENT_2_SHORT + UNSTRESSED_LONG + UNSTRESSED_SHORT + CODA_R + PHRASE_FINAL + *log*(LEXICAL_FREQ) + *log*(SPEECH_RATE) + AGE + CLASS*RACIALIZATION:NETWORK_BRIDGE _PREDICTOR + (1|SPEAKER) + (1|VOWEL)

Note that the use of three-way interactions for a sample size of 29 is not really appropriate, but it offers a framework for future expansions of this sort of network investigation. Table 9.13 contains the outputs of each model. Both Models 1 and 2 show significant interactions. Network properties in Model 1 are the most robust in terms of coefficient significance, AIC, and BIC. The implications of Model 1 – should it be an accurate representation of speech in Stockholm 1 – is best exemplified via three types of *svensk* speakers – higher-class, middle-class, lower-class – with three different types of networks: (1) has zero 'bridge' contacts in his network, (2) has eight 'bridge' contacts in his network, and (3) has 16 'bridge' contacts in his network. The range 0 to 8 to 16 is used because this is the approximate bridge-count range that actually exists in the data (Table 9.11). The base phrase used for the case study calculation is 'tidigare på dagen satt jag uttråkad' (Earlier in the day I was sitting around bored):

	Case study Higher-class *svensk*			Case study Middle-class *svensk*			Case study Lower-class *svensk*		
n Bridge3 contacts:	0	8	16	0	8	16	0	8	16
estimated nPVIV:	57.8	53.2	48.7	58.1	67.9	77.7	58.4	82.6	106.8

The calculation implies that having a high number of weak contacts among higher-class *svensk* speakers is a predictor of lower nPVIV. In contrast, the reverse effect is expected for *svensk* speakers from the lower classes.

Let us return to the question whether Stockholm's multiethnolect is responsible for lower alternation among elites. While the statistical models here are weak, they push the evidence toward the affirmative. Model 1 predicts a reduced nPVIV for the *svensk* higher class, and it would appear that this is *only* an effect for the higher classes. For highly-connected middle- and working-class *svensk* speakers, it predicts high nPVIV values. This means that either the model is inadequate or that the connectedness of working-/middle-class weak ties is different from the connectedness of higher-class weak ties. What I mean is that weak ties are said to be prevalent among higher-class speakers, which means that their weak ties are connected to other people with weak ties. On the other hand, a working-class speaker with many weak ties may be an anomaly, which means that those weak ties dead-end into dense, disconnected clusters. This would, per Bergs' (2005) model on page 264, facilitate the maintenance of conservative features.

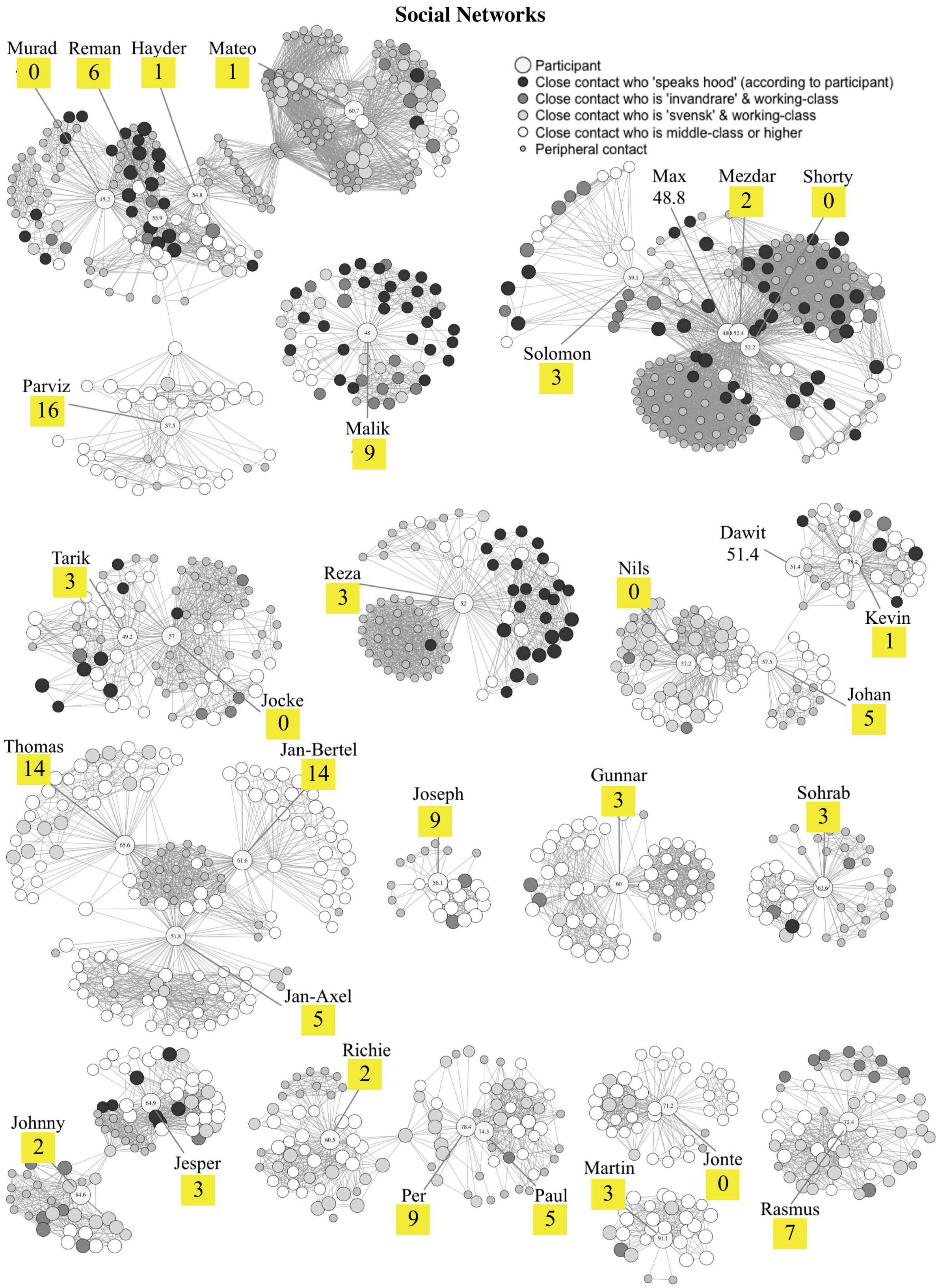


Figure 9.13: Social networks of the 29 speakers who participated in the social-network interview plus two speakers whose networks were partially mapped via participation in other ego stars (Dawit, Max). The 29 networks fall into 15 clusters. Participant nodes (yellow) are annotated with the participant's pseudonym followed by his *Network Bridges*$_3$ count (number of contacts with three or fewer mutual contacts).

	Model 1 *Network Bridges*$_3$		Model 2 *Network Bridges*$_4$	
Response variable	nPVIV (casual)		nPVIV (casual)	
(Intercept)	−16.16 (11.23)		−14.05 (11.72)	
Internal predictors		vif		vif
accent_1_long·*yes*	23.99 (1.16)***	1.2	23.99 (1.16)***	1.2
accent_1_short·*yes*	14.20 (1.33)***	1.1	14.20 (1.33)***	1.1
accent_2_long·*yes*	19.07 (1.68)***	1.1	19.06 (1.68)***	1.1
accent_2_short·*yes*	10.79 (1.40)***	1.1	10.79 (1.40)***	1.1
unstressed_long·*yes*	8.61 (1.10)***	1.3	8.61 (1.10)***	1.3
unstressed_short·*yes*	15.53 (1.23)***	1.3	15.53 (1.23)***	1.3
coda_r·*yes*	6.72 (1.55)***	1.0	6.72 (1.55)***	1.0
phrase_final·*yes*	6.11 (1.28)***	1.0	6.11 (1.28)***	1.0
log(lexical_freq)	−0.55 (0.13)***	1.0	−0.55 (0.13)***	1.0
log(speech_rate)	5.76 (0.96)***	1.1	5.77 (0.96)***	1.1
Social predictors: all models		vif		vif
age	0.81 (0.36)*	1.3	0.77 (0.36)*	1.3
class	−0.01 (0.12)	2.3	−0.02 (0.13)	2.3
Social predictors: only in Model 1		vif		vif
racialization·*svensk*:network_bridge_3	3.02 (1.17)**	2.5	-	
racialized_binary·*invandrare*:network_bridge_3	−0.52 (1.13)	2.5	-	
class:racialization·*svensk*:network_bridge_3	−0.04 (0.02)$^{\circ}$	2.9	-	
class:racialization·*invandrare*:network_bridge_3	0.01 (0.02)	2.9	-	
Social predictors: only in Model 2		vif		vif
racialization·*svensk*:network_bridge_4	-		1.96 (0.87)*	2.6
racialized_binary·*invandrare*:network_bridge_4	-		−0.71 (1.12)	2.6
class:racialization·*svensk*:network_bridge_4	-		−0.02 (0.02)	2.9
class:racialization·*invandrare*:network_bridge_4	-		0.01 (0.02)	2.9
aic	171 598.05		171 600.25	
bic	171 752.15		171 754.35	
Log Likelihood	−85 779.03		−85 780.13	
Num. obs.	16 394		16 394	
Num. groups: vowel	2 760		2 760	
Num. groups: speaker	29		29	
Var: vowel (Intercept)	75.64		75.63	
Var: speaker (Intercept)	61.62		64.36	
Var: Residual	1 989.18		1 989.19	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\circ}p < 0.1$

Table 9.13: Mixed-effects linear regression models for the effects on nPVIV of network bridges: Model 1 uses number of *bridges*$_3$, defined as contacts who only know THREE or fewer other contacts. Model 2 uses number of *bridges*$_4$, defined as contacts who only know FOUR or fewer other contacts. nPVIV in CASUAL speech is the response variable; all models contain the same internal predictors and the same social predictors: age, racialization, and social class (defined as the first principle component *PC1* of SEI, income, education, parental SEI, parental education, and *Mosaic* taste). For categorical predictors, the reference category is in italics. Coefficients are indicated in the center column, standard errors in the parentheses, and variance inflation factors (VIF) to the right.

6 Chapter summary

This chapter sought to address the question of apparent-time stratification of rhythm in Stockholm and its evolution. The evolution of rhythm in Stockholm, according to the four analyses conducted, can be summarized as follows:

Analysis 1 is visualized in Figure 9.1 (p. 236). Rhythm among older working-class *invandrare* is relatively staccato compared to other speakers of the same age. Yet, it is not stylistically sensitive. As apparent time progresses, the speech of working-class *invandrare* becomes more and more staccato while also becoming more stylistically sensitive. The conclusion drawn from this is that staccato rhythm has recently moved from indicator to marker status, implying that it is a relatively novel feature. Meanwhile, the speech rhythm of working-class *svensk* speakers appears to be highly non-staccato, and this feature appears to be established with no change in stylistic sensitivity. The conclusion drawn from this is that non-staccato rhythm is a legacy feature of Stockholm's traditional working-class Södersnack. A final interpretation from this analysis is that rhythm is generally becoming more staccato over apparent time in Stockholm, independent of social factors.

Analysis 2 is best conceptualized by examining Figure 9.3 (p. 239) in conjunction with Figures 9.4 (p. 240) and 9.5 (p. 241). A change-point exists for rhythm between those born before and after 1983. This change point corresponds with three transformative macrosocial moments in late-modern Swedish history. The first is the second spike in refugee migration that occurred in the mid 1990s, illustrated in Figure 9.4. This spike coincides with the onset of adolescence for the cohort born after 1983. The second historical event was the 2001 tipping point in school segregation identified by Holmlund et al. (2014), illustrated in Figure 9.5 (p. 241). In examining the year-specific demographics of the schools that the participants in *this study* attended, the same change-point in segregation exists between 2000 and 2001. The third historical event was the spike in income inequality in year 2000. When the evidence from Analysis 2 is considered alongside the fact that school diversity is highly correlated with nPVIV (recall Model 9 in Chapter 6, p. 173), it can be concluded that transformative migration, coalescing with sudden new stratifications and enclosures around adolescents, has likely been a contributing factor to the expansion of staccato rhythm.

Analysis 3 showed, if we are to accept the apparent-time premise, that the variation in rhythm has not evolved as a unitary construction. In the early 1980s, working- and middle-class *invandrare* began reducing accented long vowels and enlarging accented short vowels. In the mid 1980s and early 1990s, working- and middle-class *invandrare* partly restored the size of accented vowels. They then initiated the enlargement of unstressed vowels in the domains of duration, intensity, and f_0. Presumably, this has resulted in the hyper-staccato effect characteristic of this younger cohort.

Among *svensk* speakers, much of the class-based rhythmic variation also lies within vowel pairs that only contain unstressed syllables. However, the direction by which it stratifies is split between two generations. For unstressed vowels next to unstressed vowels, class has a positive relationship with the staccato effect among the *svensk* speakers born after 1976. Class has an inverse relationship with nPVIV among *svensk* speakers born in 1976 or earlier (p. 249).

Analysis 4 investigated whether connectedness to the greater speech community could be responsible for the lower nPVIV of the *svensk* higher class. The data available was insufficient to make any definitive conclusions. However, connectedness by means of 'weak bridges' did have some predictive power for nPVIV. Highly-connected higher-class *svensk* speakers are predicted to have lower nPVIV scores than poorly-connected higher-class *svensk* speakers. The reverse is predicted for lower- and middle-class *svensk* speakers, which could be due to the fact that highly-connected speakers from the working classes are merely anomalous conduits to other non-conduited clusters. In contrast, highly-connected speakers from the upper classes are more the norm and are therefore conduits to other conduits, making them more connected to diffusing features.

Citations in original language

[i] Original Swedish conversation between Mezdar, Max, Shorty:

Shorty: å de känns som att förortssvenska är mer kärleksfullt
Mezdar: Ah {yawn}
Shorty: Asså det mer sär familjärt
Mezdar: Man kan visa mer kärlek bror
Shorty: Alla e bröder, alla e systrar, alla e- fattar du
Mezdar: Ah, den e inte kall
Shorty: Nä
Mezdar: Men bror haru tänkt på en grej? Dom som e äldre i Solomons eh generation exempelvis
Shorty: Ah, ah ta ((nittiosjua, åttor))
Mezdar: Solomon, eh han som äger Cafe Mynta
Shorty: Berra
Max: Han pratar ren *svensk*a
Shorty: Ah
Mezdar: In- deras *svensk*a, den e inte eh ren. Den e lite eh...
Shorty: Ah ja ja
Mezdar: Den e m- mindre än
Max: Den e tydligare och den är ehm
Mezdar: Tydligare å dom har
Max: Det inte så mycket sjukt slang om man säger så
Mezdar: Nej de låter mer som eh *svensk*a
Max: De finns ingen, ah det {laugh} ingen idé alls
Shorty: Exakt exakt
Mezdar: Jämfört me oss den e fett ren, men jämfört me en *svensk* man märker att den här killen
Shorty: Han e lite
Mezdar: Han har vart me om saker å ting {laugh} förstår du

Shorty: Ah ah ah ah ah {laugh} ah han e lite förort
Mezdar: Ah å sen vi, det lite mer. Det har kommit nya ord, påhittade ord
Shorty: Ja ja ja ja ja ja ja
Mezdar: Använda ord vi har n- ord från andra språk
Shorty: Exakt slår ihop två språk
Mezdar: Men bror, nu dom härа y- yngre än oss, bror, ja förstår inte ens va dom säger
Shorty: Vi hinner inte!

Chapter 10

SUMMARY OF FINDINGS

The four analysis Chapters 6, 7, 8, and 9, while at times creeping into other domains of investigation, each spring from a series of interconnected research questions.

> RHYTHM IN THE VERNACULAR: *Does the speech of Stockholm's racialized working class have lower rhythmic alternation than other varieties in the city; i.e., is it 'staccato'?*
>
> PHONOLOGY OF RHYTHM: *Is the variation in rhythm caused by (a) specific change(s) in the vowel system of certain speakers?*
>
> RHYTHM IN CONTEXTUAL STYLE: *To what degree is speech rhythm stylistically sensitive in Stockholm, and what inference can be made about the maturity of its associated variants?*
>
> THE EVOLUTION OF RHYTHM: *What is the history of rhythmic stratification in Stockholm, and how has it evolved?*

The key findings from this dissertation are presented here.

1 Rhythm in the vernacular

1. Male working-class *invandrare* speakers are predicted to produce the lowest nPVIV values (*most staccato*) in their vernacular speech. The same trend holds for speakers who attended schools with high numbers of students of foreign descent. The same

trend also holds for speakers who have grown up in neighborhoods with a high number of residents of foreign descent.

2. Network qualities that pertain to the metapragmatics of speaking 'hood' correlate with lower nPVIV values (*more staccato*), which further strengthens the *staccato* evaluation of multiethnolect.

3. Male working-class *svensk* speakers are predicted to produce the highest nPVIV values (*less staccato*) in their vernacular speech. This is to say that the interaction of *svensk* and being working-class predicts an opposite effect on nPVIV to the interaction of *invandrare* and being working-class: higher nPVIV for the former and lower nPVIV for the latter.

4. As social class increases, racialization loses its predictive power. Higher-class speakers are predicted to have intermediate-level nPVIV values in their vernacular speech, regardless of racialization.

5. Younger speakers are predicted to have generally lower nPVIV values (*more staccato*) than older speakers, independent of social factors and contextual style.

6. The class-related predictor with the strongest stratifying effect on nPVIV is parental SEI, which implies the primacy of embodied class on language practice versus measurements of less embodied class (i.e., SEI, income, formal education).

7. Network qualities have qualitative explanatory power for individual speakers whose nPVIV contradicts the traditional social-class trend. For example, a speaker like Mateo who is *invandrare* and working class, yet who produces a high nPVIV, will likely have many close contacts who are *svensk* and working-class. A speaker like Shorty who is *invandrare* and middle class, yet who produces a low nPVIV, will likely have many close contacts who are *invandrare* and working-class.

The evidence presented in this section supports an affirmative answer to the Research Question: *Does the speech of Stockholm's racialized working class have lower rhythmic alternation than other varieties in the city; i.e., is it 'staccato'?*. The analysis shows that for men, racialized and working-class identities – according to almost all class parameters – will predict lower nPVIV (*more staccato*) in CASUAL speech. Rhythmic alternation is defined as the alternation of strong and weak elements, and it is measured with normalized pairwise variability index of vowels (nPVIV) using the energy-f_0-integral (EFI) from each vowel.

2 Phonology of rhythm

1. The multiethnolectal patterns in nPVIV – that is, the nPVIV of the lower and lower-middle class *invandrare* – appear to be a function of the shortening of prominent long

vowels and the lengthening of non-prominent short vowels. In contrast, the shortening of non-prominent long vowels, and the lengthening of prominent short vowels appear to play less of a role behind the social stratification of rhythm. The former is socially stratified but has no connection to nPVIV (compare matrix with Figure 7.2's Model 3); the latter has a connection to nPVIV but is only socially stratified between the lower-class *invandrare* and the lower-class *svensk* (compare matrix with Figure 7.2's Model 2).

2. Lower-class *invandrare* lengthen phrase-final vowels by 16.2 milliseconds more than lower-class *svensk* men, 9.1 milliseconds more than high-class *svensk* men, and 4.6 milliseconds more than higher-class *invandrare* men. As it concerns the question of staccato rhythm, variation in phrase-final lengthening is not contributing to lower alternation in multiethnolect. Rather, it appears to be its own separate variant.

3. No categorical monopthongization is underway for lower- or upper-class *invandrare*. Rather, it appears that staccato rhythm may be coercing some monophthongization in the case of prominent vowels and permitting fuller diphthongal realizations in the case of non-prominent vowels.

4. *Invandrare* speakers have a larger vowel space area (VSA) for all vowels: prominent long and short and non-prominent long and short. For long vowels, they appear to be leading in phonological change, which is implied by the clear lowering chain shift in Figure 7.9, which is rendering the VSA larger. In the case of non-prominent vowels, the findings show that *invandrare* are reducing less than *svensk* speakers. This result implies that *svensk* speakers having more schwa-like unstressed vowels and that *invandrare* speakers have fuller realizations of unstressed vowels (Table 7.3).

The evidence presented in this section supports an ambiguous answer to the Research Question: *Is the variation in rhythm caused by (a) specific change(s) in the vowel system of certain speakers?*. One key component to the reduced alternation in multiethnolect is the shortening of prominent long vowels and the lengthening of non-prominent short vowels. Other transformations appear to be independent of rhythm – such as phrase-final lengthening – or a function of rhythmic coercion, such as the monophthongization of LETA.

3 Rhythm in contextual style

1. In formal styles, working-class *invandrare* and working-class *svensk* speakers attenuate their production of staccato and non-staccato speech, respectively. The rhythmic pattern that they produce in formal speech appears to target the casual speech produced by higher-class speakers.

2. The majority of social predictors and social-network predictors have significant predictive power for the variation of speech rhythm in casual speech. In more formal styles, most of these predictors lose predictive power.

3. Younger speakers, regardless of social background, are predicted to have more staccato speech than older. This is stylistically robust – the main effect of age is only mildly weaker in more formal styles

Returning to the Research Question *To what degree is speech rhythm stylistically sensitive in Stockholm, and what inference can be made about the maturity of its associated variants?*, the evidence points to rhythmic stratification for both working-class groups being quite salient. This implies that both the staccato variant of lower-class *invandrare* and the non-staccato variant of lower-class *svensk* have some degree of maturity.

As it pertains to the gradient age distribution of rhythm across all speakers, regardless of class and racialization, there does not appear to be stylistic sensitivity. In other words, the slope for age is constant across all contextual styles (Figure 8.4, p. 223). Therefore, it can be concluded that the more intense social stratification of rhythm among the lower classes is salient but that the mild generational stratification is not.

4 The evolution of rhythm

1. As apparent time progresses, the speech of lower-class *invandrare* men is becoming more and more staccato while also becoming more stylistically sensitive (visualized on p. 236). The conclusion drawn from this is that staccato rhythm has moved from indicator to marker status.

2. The speech rhythm of lower-class *svensk* men appears to be highly non-staccato, and as apparent time progresses, this rhythmic pattern maintains a consistently high degree of stylistic sensitivity. The conclusion drawn from this is that non-staccato rhythm is a legacy feature of Stockholm's traditional working-class Södersnack.

3. The variation in rhythm has not evolved uniformly as a unitary construction. Rather, lower- and middle-class *invandrare* in the early 1980s began reducing accent-1 and accent-2 long vowels and enlarging accent-1 short vowels. Starting in the mid 1980s and early 1990s, lower- and middle-class *invandrare* partly restored the size of accent-1 long vowels while continuing the reduction of accent-2 long vowels and the enlargement of accent-1 short vowels. They then initiated their own innovation, which was the enlargement of unstressed vowels in the domains of duration, intensity, and f_0. Presumably, this has resulted in the staccato effect.

4. The apparent-time change point in rhythm that is described above corresponds with three transformative macrosocial moments in late-modern Swedish history – a sud-

den upswing in migration, a peak in income stratification, and a sudden increase in school segregation. I conclude that transformative migration, coalescing with sudden new social and spatial enclosures around adolescents, has likely been pivotal to this secondary innovation for staccato rhythm.

5. I will tentatively propose that social-network connectedness by means of 'weak bridges' has some predictive power for low nPVIV among higher-class *svensk* speakers. Highly-connected higher-class *svensk* speakers are predicted to have lower nPVIV scores than poorly-connected. The reverse is predicted for lower- and middle-class *svensk* speakers, which could be due to the fact that highly-connected speakers from lower classes are merely anomalous conduits to other non-conduited clusters. Meanwhile, highly-connected speakers from the higher classes are more the norm and are therefore conduits to other conduits, making them more connected to diffusing features.

Returning to the Research Question *What is the history of rhythmic stratification in Stockholm, and how has it evolved?*, the answer is that rhythm in multiethnolect has evolved in two main stages. The older cohort that grew up in more traditionally 'modern' social conditions have significantly reduced accented vowels while the remainder of their phonology is relatively normative (as it pertains to higher social groups). The younger cohort that grew up in hyper-segregated 'late-modern' social conditions have significantly reduced unstressed vowels, rendering a phonological inventory that is almost entirely non-normative (as it pertains to higher social groups). Overall, however, speech rhythm is changing regardless of social category. A small and incremental 'staccato-ization' is diffusing into mainstream Stockholm Swedish, and preliminary evidence identifies multiethnolect as potential source.

—

The findings here add weight to the observations that Swedish multiethnolect is 'staccato' – in as much as one accepts the connection between racialized working-class speakers and this so-called variety. The findings also support earlier observations in the literature that certain speakers reduce the staccato effect in more formal styles. Interesting also is the finding that staccato rhythm has become more stylistically sensitive in apparent time. This implies that it is a relatively new rhythmic variant.

Turning to the working-class *svensk* speakers, the findings are somewhat surprising and imply that rhythm has been a stratified variable long before the emergence of contact prosody in Stockholm's suburbs. The speech rhythm of working-class *svensk* speakers diverges more from the speech rhythm of higher-class speakers than the speech rhythm of working-class *invandrare* speakers diverges from higher-class rhythm. The high alternation of the *svensk* working class also shows stable stylistic sensitivity in apparent time. Therefore, the possibility arises that rhythm might be an old sociolinguistic variable in Stockholm.

If high alternation indeed was an original feature of *Södersnack*, then the social stratification of rhythm predates even the migration waves of the 1970s. When one considers this possibility in tandem with the tension between the two working classes, one can envision a relatively permissive social ecology for the selection of staccato rhythm when the contact feature pool first emerged. As Cheshire et al. (2011) have argued, salience is a key determiner of feature selection.

The findings also show that the stratification of rhythm among working-class *invandrare* cannot be ascribed to one segmental feature such as the reduction of accent-1 vowels. Instead, the findings identify contrast between successive vowels as a governing framework. The apparent-time adjustments of duration, for example, have not been unidirectional on phonologically short or long vowels. Rather, they appear to be situational, going in the direction that results in a lower rhythmic contrast.

Missing, however, from this analysis is a perceptual study. While this study has shown that alternation varies in production according to prescribed definitions of rhythm, the link between this and perception is an open question. For example, do listeners actually perceive the percentage differences that nPVIV measures, or are they responding to, say, the absence of schwa (pp. 200, 205)? My hope is that this study will set into motion analyses that address these types of follow-up questions. In the next chapter, I discuss the contributions I have hoped to make with this study along with its wider relevance to sociological and sociolinguistic theory.

Chapter 11

DISCUSSION

1 Methodological contributions

This dissertation makes methodological contributions to the analysis of rhythm and phonetics, and it offers a path for exploiting the expansive data offered by Nordic population registers for the purpose of enriching sociolinguistic analyses.

Rhythm and phonetic investigations

As I described on page 131, up until now, investigations of rhythm have based their analyses on median (or mean) calculations for each speaker. This is not just the case for the studies that have used nPVIV, but also studies that have used metrics like variance coefficients or standard deviations. The approach that I take in this dissertation is to treat every pair-level nPVIV calculation as a single observation while simultaneously coding for internal constraints. While new for rhythm analysis, this approach actually moves the sociolinguistic investigation of rhythm closer to the established variationist approach: accounting for ALL constraints – internal and external – when attempting to account for variation. This dissertation offers a detailed roadmap of how to do that.

This project was also the first to make use of 'big data' in the analysis of Swedish in Stockholm. This was facilitated by the development of SweFA, a forced aligner for Swedish, described on pages 77–86. It made possible the processing of eight times the amount of data than otherwise would have been possible for a project of this scope. Due to constraints on space, I have not provided all the details of how this aligner was developed. However, I outline the steps that differ from the guidelines in circulation and are unique to this project. These are the creation of a prototype from the American hidden Markov models in FAVE and the workaround for training the Montreal Forced Aligner on manually-corrected Textgrids.

Using statistics from Nordic population registers

As Kerswill and Williams (2000) remind us, investigations of dialect formation are often post-hoc, "in that they examine the outcomes of the process several generations after the original migration" (2000). This study is no exception; it relies on the connection between historical data and linguistic patterns in apparent time. This makes the reliability of such historical data all the more acute.

The Nordic countries are unique because they keep detailed population registers and offer robust statistics that are based on the full citizenry. In contrast, countries like the US or UK produce statistics that are the results of generalizations made from intermittent census information. School enrollment data is even less reliable in these countries because the numbers come from the schools themselves, are sifted through local-government bureaucracy, and then reported to either a trust or a national-government agency. In the Nordic countries, a central statistics agency links every citizen's national ID number to nearly every instance of institutional enrollment and membership (in Sweden, even your gym enrollment is tied to your ID number). The agency then produces 100%-accurate demographic data on institutions like schools. It is therefore the case, as I hope this study has shown, that investigations of Nordic speech communities can be especially potent in testing sociolinguistic theory.

This study has not even fully realized the potential of the data available. I'll offer an example. I collected publicly-available information that includes one crude social-class parameter and one crude ethnicity parameter for school, school track, and year. Forsberg (2018), however, has purchased demographic data on every secondary-school pupil in Sweden that ties school, school class, and school year to every pupil's parental income, profession, education level, country of birth, postcode, and so on. Naturally, this level of detail is expensive and can only be purchased from Statistics Sweden with research-institutional clearance. However, with the appropriate funding, future sociolinguistic investigations could to take advantage of this sort data. It would facilitate constructing a more detailed speaker profile than has ever been possible before. From there, the questions seem limitless. What classroom constitutions seem to be the most conducive to certain apparent-time differences? Did speakers who style-shift the most now have more middle-class kindergarten classmates? Or does it correspond most closely to the classmate profile of their middle school? Or is their no correlation at all, leading us to focus on identity politics and language aptitude?

2 Contributions to traditional Swedish dialectology

Any analysis of variation and change is of importance to the field of dialectology. However, the presence of a large number of L2 speakers in a community challenges traditional notions about what dialectology 'cares about'. Even today in some circles, there remains a sentiment that the changes led by multiethnic youth in the immigrant periphery are not of dialectological concern, better left to researchers of bilingualism and language acquisition.

It is not my goal to offer arguments against this position. Boyd and Fraurud (2010); Cheshire et al. (2011); Eliaso Magnusson and Stroud (2012); Fraurud and Boyd (2011); Quist (2008); Stroud (2004) and others have all offered compelling reasons to include L2 speakers in traditional dialectology. Boyd and Fraurud (2010) even offer evidence to the fact that 'L2' as a notion is highly-flawed in contemporary urban contexts. Nonetheless, Swedish dialectology remains unique in what appears to be an institutional division between bilingualism research on one hand and dialect investigation on the other. Naturally, researchers should focus on whatever they like and whatever they can get funding for, even if it means overlooking important groups. It is safe to say that there is such a dearth of sociophonetic research in Sweden that we ought to be very supportive of any epistemological angle. With that said, I would like to identify some dialectological contributions made by this dissertation that do not beg the role of L2 speakers. In doing so, I hope to strengthen the bridge between bilingualism research and traditional dialectology.

The first contribution is the documentation of rhythm for Stockholm Swedish. Whereas traditional dialectologists may search for speakers considered iconically 'authentic', I have sought out those considered iconically non-authentic[1]. But a descriptive account of minority speech is meaningless without an account of majority speech for comparison. Therefore, I have also done 'traditional' dialectological work by describing rhythm for more iconic Stockholmers like the white *svensk* working class. I show that among these white *svensk* speakers in Stockholm, social class plays a potentially greater role in rhythmic variation than among the *invandrare* speakers. I also show that the speakers with the highest iconic link to Stockholm, namely, this established *svensk* working class, show some of the highest rhythmic differences of any group from the central rhythmic tendencies of the wider community. Academic discussions about rhythm in Stockholm typically lead to the topic of L2 speakers; however, the data here show that such discussions should just as readily lead to the topic of the Söder diaspora (see p. 35 regarding this term).

The second contribution that this analysis makes is it documents robust participation in the Central Swedish Vowel shift in Stockholm. It shows which vowels appear the most advanced and which groups appear to be leading (p. 192). In examining this result, it becomes impossible to ignore the role of so-called 'L2 speakers'. Both the lower-class and higher-class *invandrare* are exceptionally advanced. As I discuss on page 1.1, Nordberg (1975) and

[1]In fact, this is not a Stockholmian phenomenon. Racialized speakers are often denied the association of any regionality at all; see, e.g., Cornips (2019, forthcoming)

Leinonen (2010) report that a downward chain shift has been occurring in Central Swedish long vowels for some time. Nordberg (1975) found that speakers in the industrial town of Eskilstuna produced the SÖT vowel like the more open vowel in DÖR. This 'hypercorrection', as he framed it, likely pulled LUS in such a manner that it resembles the old SÖT today. He also found that young speakers were producing NÄT like open LÄR and that this was stratified between young and old speakers. Leinonen (2010) found this same stratification in her study between old and young speakers for both *Standard Swedish* and *speakers in Stockholm* for the NÄT-to-LÄR and SÖT-to-DÖR admergers (2010:162–163). What is more, she found a consistent lowering for ALL front long vowels among young speakers throughout Sweden. Gross et al. (2016) also identified a robust NÄT-to-LÄR merger among youth in Stockholm. Their finding, however, showed this change to be socially-neutral, which is not necessarily contradictory to my findings. The participants in my study average a good 10-15 years older than the participants in Gross et al. (2016).

The results here imply that if one were to only include 'authentic' speakers when investigating the Central Swedish Vowel in Stockholm, one would be left empty-handed. It appears that Stockholm's racialized working class is leading this particular phonological advancement. It is not the first time that 'newcomers' have been found to be the most advanced in shifts that were otherwise considered to be highly indigenous. In their investigation of the Canadian Vowel Shift in Toronto, Hoffman and Walker (2010) found that Italian Canadians produced the most advanced variants. Similarly, Hall-Lew (2010) found that the Asian Americans in her study of San Fransisco English were the most advanced in a number of variants associated with the California Vowel Shift. Both studies, along with the findings presented here, demonstrate that traditional dialectology ignores nontraditional speakers at its peril.

3 The epiphenomenality of rhythm

One goal of Chapter 7 was to test whether it was appropriate to treat rhythm as its own variable or not. If nPVIV was merely capturing some other single extraneous segmental variable, such as the durational shortening of phonologically long vowels or the monophthongization of diphthongal vowel phonemes, this would be revealed. If, on the other hand, nPVIV was influenced by a number of simultaneous segmental transformations, some going in opposite directions, this would strengthen the case for a unitary concept of rhythm. In that sense, investigating the phonology behind the variation in nPVIV would 'kill two birds with one stone'.

The evidence from Chapter 7 points toward the latter conclusion, especially when the results of Chapter 9 are also considered. Assuming that the speech rhythm of Stockholm's lower-class *invandrare* was being caused by the systematic shortening of long vowels or the systematic lengthening of short vowels, one would expect this transformation to occur in both acoustically prominent and non-prominent position. The evidence from Chapter 7 identifies only acoustically prominent long vowels as undergoing shortening and acoustically non-prominent short vowels undergoing lengthening. This is when *invandrare* speakers are divided at 50 for the first principle component of social class. The evidence also shows that no consistent monophthongization is occurring for diphthongal long vowels except for LETA.

Turning to Chapter 9, the transformations are also not uniform. Unstressed long vowels and unstressed short vowels appear to undergo the same transformation of enlargement (p. 249). It is only when they are either pretonic or posttonic that the model sees distinctions in phonological duration (p. 250, 251, 252).

Recall in my review of approaches to rhythm (p. 102) that Bruce (1983) proposed that Swedish rhythm might operate epiphenomenally, beyond even the dynamic of prominence. In his pilot study he found that the unstressed syllables between two stressed syllables in a prosodic phrase fell into an alternating pattern of *weak-strong-weak*. He offers the following explanation.

> It is my bias in thinking about rhythm in spoken language that the leading principle is alternation and not isochrony. [...] The principle of rhythmic alternation entails a conspiracy against sequences of several equally prominent syllables. This applies to both unstressed and stressed syllables. (Bruce 1983:36–37)

The findings here have not investigated the existence of a strong-weak-strong pattern between accents that Bruce found. Nonetheless, they have implied the existence of such a 'conspiracy'. Instead of coercing the rhythmic grid into low alternation, phonetic transformations seem to be coerced by the effects of low alternation.

This has interesting implications in light of the literature to date on the role of rhythm in language acquisition (Ramus et al. 1999) and language processing (Cutler, Mehler, Norris, & Segui 1986; Kolinsky, Morais, & Cluytens 1995; Mattys & Melhorn 2005; Nazzi, Bertoncini, & Mehler 1998; Nazzi, Iakimova, Bertoncini, Frédonie, & Alcantara 2006). This literature has implied that rhythmic features can be prior and separate from their segmental components, strengthening the interpretation here that reduced contrastive rhythm, i.e., 'staccato', is a linguistic construct in its own right.

4 Two working classes with opposing rhythmic variants

An unexpected finding in this study was that Stockholm's two working classes, one *svensk* and one *invandrare*, have two linguistic features that are diametrically opposed to one another. These include speech rhythm as measured by nPVIV and phrase-final lengthening. These can be summarized as follows:

	invandrare lower class	higher class	*svensk* lower class
mean nPVIV (p. 155)	53.5	⟵ *61.3* ⟶	69.3
mean duration phrase-final vowel (p. 188)	109.1	⟵ *100.9* ⟶	92.1

What is particularly interesting about these two variables is that they both constitute elements of rhythm; yet, they are moving in opposite directions. The staccato effect of a low nPVIV would be enhanced by less phrase-final lengthening just like the non-staccato effect of a high nPVIV would be enhanced by more phrase-final lengthening. Instead, however, opposite trends coexist.

It is difficult to explain these differences from demographic data alone. The *svensk* working class lives in more ethnically diverse neighborhoods than any other *svensk* group (recall the lifetime neighborhood diversity of this group in Table 6.15, p. 171). Furthermore, in the so-called 'golden era' of social democratic Sweden, it flourished in the very same Million Houses estates as newly-arriving migrants from non-Western countries.

Therefore, I can think of two possible explanations for why these two features are so different. The first is identity-political. The threat of an imposing non-white working class has led to the cultivation of iconically opposed features. The second is that their social networks are so homogeneous that the surrounding multietnolectal innovations are either negatively sanctioned by ringleaders or never make it into their networks. These explanations are by no means mutually exclusive. They can both be simultaneously true.

There is some evidence to support each of the above explanations. As it pertains to identity politics, four of the *svensk* working-class participants voted Sweden Democrat in the last election. Even the ones with more progressive political leanings expressed a feeling of tension with the surrounding migrant population. This was especially the case for Jesper, a young speaker with very salient Södersnack features. Raised in Botkyrka, near where Open Society Foundations (2014) conducted its study (see review pp. 35–37), his lifetime neighborhood diversity is 81 (p. 171). He told stories of being bullied for being one of the only white Swedes in his school. His network, however, is complex. His friends are mostly ethnic Swedes, his colleagues mostly of migrant origin, and he has a half sister who is biracial[2] and speaks the typical 'ghetto language' of Botkyrka.

As it pertains to social networks, the *svensk* working class indeed has some of the most homogeneous and closed networks. Among the five working-class speakers who participated in the network interview, the percentage of close contacts who were *svensk* and working-class ranged between 39% for Jesper and 92% for Per (p. 171).

The literature has not explicitly addressed whether late-modern superdiversity has motivated any explicit linguistic avoidance among Europe's white working classes, but it has been implied. Torgersen and Szakay (2012) found that the nPVIV of young white speakers differed the most from young non-white speakers in Hackney. Older and very old white speakers in Hackney had intermediate levels of nPVIV (2012:829). This could imply a distinctive move among younger white speakers who might be targeting higher-alternation varieties in Essex

[2] He used the term 'mulatto' in the interview and did not specify the nationality of her father.

rather than the new local vernacular. Gates' (2018) analysis of speech practices in an East London school showed similar results. While nearly all students in her dataset were producing monophthongal FACE and PRICE variants, white girls were using diphthongal variants. She suggests that "they linguistically and socially construct Whiteness in a way not often found in multicultural, ethnically-diverse communities" (2018:47). Conflict models of change have of course been explored elsewhere, and the topic is a signature subject for variationist sociolinguistics. In his seminal study of local fisherman on Martha's Vineyard, Labov (1963) found that "high centralization of /ai/ and /au/ [was] closely correlated with expressions of strong resistance to the incursions of the summer people", i.e., outside vacationers (1963:297).

As I reviewed on pages 35–37, Watt et al. (2014) conceptualize white working-class displacement within a diasporic framework.

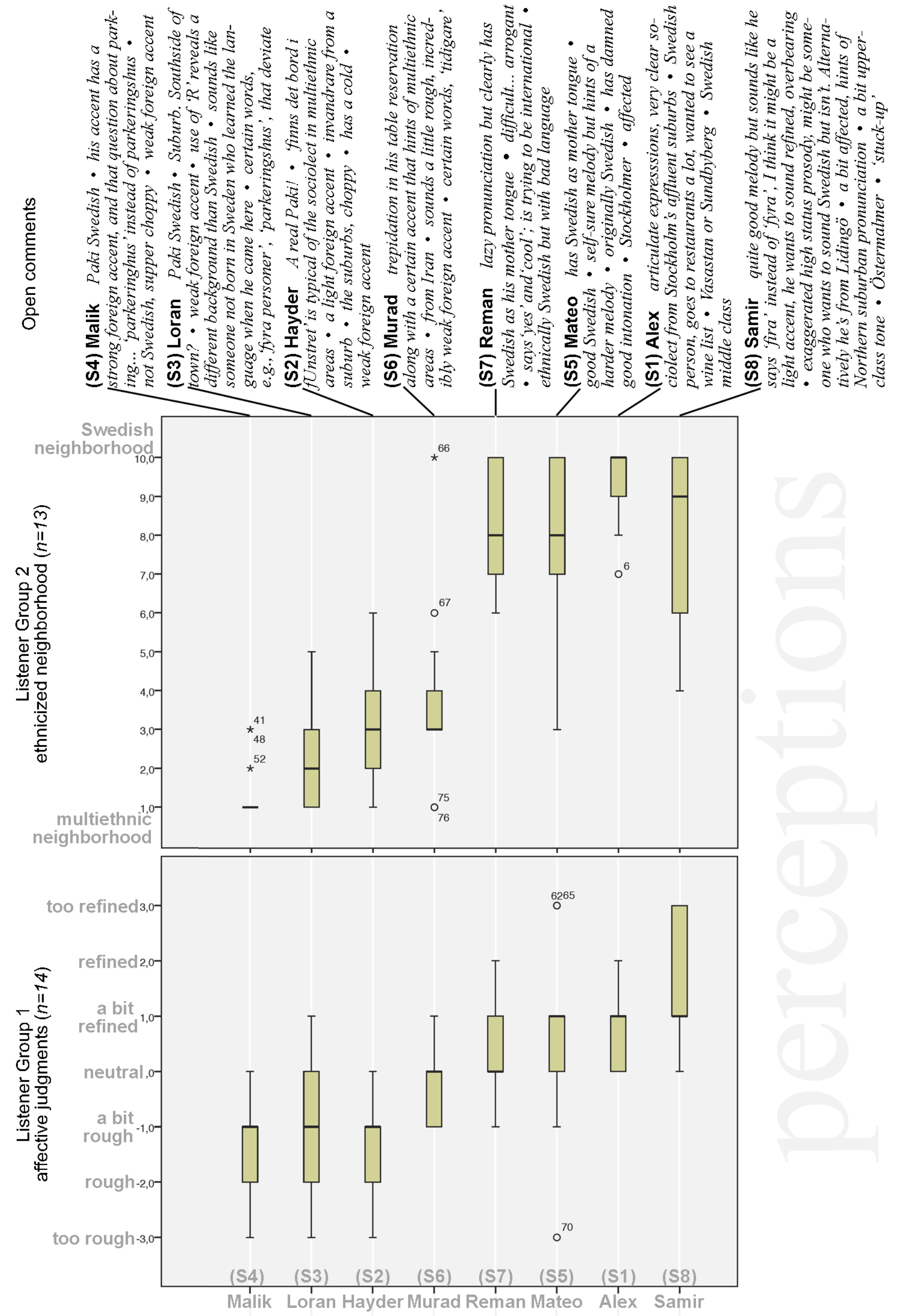
> Although this is not a diaspora proper in the generally accepted sense of constituting a transnational flow of migrants across national borders, it nevertheless has certain diasporic qualities in relation to notions of dispersal, scattering and being in one place but identifying with another (Watt et al. 2014:121).

This framework is helpful for understanding the linguistic results found here. The Million Houses Program coincided with rapid de-industrialization, the gentrification of Södermalm, and the dispersal of a homogeneous working class into ethnically diverse 'New Suburbs' like Rinkeby and Flemingsberg. With the progression of late modernity, these communities have diversified even more rapidly. At the same time, racializing forces from outside and within have contributed to the fusing of these ethnicities into a monolithic *invandrare* identity. This has perhaps led to a more acute focus on whiteness and Swedishness within the Söder diaspora. This salience, combined with the aforementioned dynamics of de-industrialization and displacement, may also be a reason why white working-class speakers are the most divergent in speech rhythm from the rest of Stockholm.

5 Lower-class *invandrare* and 'multiethnolect'

It may seem premature to draw an automatic link between lower-class *invandrare* and multiethnolect as a coherent variety. This is especially the case for a study that has focused on rhythm production and has not investigated other co-occurring phonetic features. However, my earlier work on perceptions offers a link between the results presented here and multiethnolect as a variety.

In Young (2014, 2018c), I explored the link between the perception of multiethnolect and specific speech properties. In this experiment, I recorded eight working-class participants while they, per telephone, reserved a table at an upscale restaurant in Stockholm. This was done to capture their more formal repertoires. Five of the participants in that study also participated in the current study: Hayder, Malik, Murad, Mateo, and Reman.



	Malik	Loran	Hayder	Murad	Reman	Mateo	Alex	Samir
(speech rhythm) nPVI-V:	39,2	40,6	45,3	47,3	52,6	52,7	54,8	54,9
(r) % apicals:	100	85	100	67	22	88	44	71
(yː) mean F2:	345	364	267	-	366	345	-	429
(salient norm violations):	0	1	1	0	1	1	1	1

Figure 11.1: Listener assessments about eight speakers of Stockholm Swedish, reproduced and translated from Young (2018c:50). Production features identified in the speaker stimuli are identified at the bottom, including nPVI-V of *duration*. Above that are affective judgments made by Listener group 1. Above that are ethnic neighborhood placements made by Listener 2. Comments are provided at the top with each listener separated by a bullet. Participants Hayder, Malik, Murad, Mateo, and Reman also partook in this study.

pseud.	profession	PC1 Social Class	mean nPVIV	median nPVIV	Variety designation in Young (2018)
Reman	social worker	0.0	54.0	46.2	standard
Malik	cook at cafeteria, street food	1.1	47.5	47.4	multiethnolect
Max	youth worker	4.4	47.2	45.4	
Antonio	caretaker, ex drug dealer	7.6	42.5	42.3	
Hayder	caretaker, bouncer	19.0	56.0	36.4	multiethnolect
Dawit	group home assistant	19.8	47.8	46.6	
Solomon	youth worker	20.1	56.2	51.9	
Murad	group manager, furniture retail	21.7	43.9	44.6	multiethnolect
Abel	photographer, personal trainer	25.4	50.3	37.7	
Mateo	waiter, bartender	31.6	59.4	42.8	standard
Reza	waiter, napropathy student	34.8	49.4	43.0	
Mezdar	youth worker (certified)	37.5	51.2	43.5	
Jocke	athletic coach, personal trainer	42.2	53.8	43.6	
Sohrab	web architect	48.6	62.8	49.4	
Shorty	social worker	51.2	51.6	45.0	
Kevin	organisational consultant	64.0	56.7	48.5	
Tarik	field assistant; athletic coach	73.4	49.1	46.1	
Majeed	doctor	76.2	64.4	52.4	
Parviz	business owner, physician	91.1	59.5	40.4	

Table 11.1: Median nPVIV values per speaker of the four largest vocalic-pair types. Speaker pseudonym, profession, and social class are listed along with how the speaker was grouped in the ctree analysis. Total medians are calculated as well in the far right column. 'Anomalous' nPVIV values are shaded in gray. The perceptual assessments of the formal register of five speakers is provided in the far right column (from Young 2018c.

I then constructed eight short stimuli of 30 seconds each from each of the eight recordings and played them to two listener groups who were asked to assess the stimuli by means of Likert scales. I also asked them to write in comments about what they heard. The details of the methodology are provided in Young (2014:14–18) and Young (2018c:46–48).

The assessments of these speakers correlated highly with their nPVIV of duration. Figure 11.1 contains a reproduction of the Likert results for Listener Group 1, Listener Group 2, and the written comments. Speakers with a lower mean nPVIV were overwhelmingly assessed by Listener Group 1 as 'rough' and by Listener Group 2 as belonging to a 'multiethnic neighborhood'. Speakers with higher mean nPVIV were overwhelmingly assessed as by Listener Group 1 as 'neutral' or 'refined' and by Listener Group 2 as belonging to a 'Swedish neighborhood'. Speakers with low nPVIV values were referred to as 'accented' and 'choppy' in the written comments.

Assessing all the perceptions together, I concluded that Hayder and Malik were producing some version multiethnolect, Murad was producing a version of 'multiethnolect light', and Reman and Mateo were speaking standard Stockholm Swedish. Since these recordings were made on formal speech, we can assume that the CASUAL-speech recordings for the same speakers would contain at least the same densities of 'multiethnolectal' features (if not more).

Given the evidence presented in Young (2018c) that speech rhythm was the main driver of these perceptions, the three speakers assessed as 'multiethnolect' can serve as anchors for the production data in this study. Table 11.1 lists the *invandrare* speakers with their respective mean and median nPVIV values. A final column has been added with designation of what variety that speaker sounds like in formal speech, according to Young (2014, 2018c). The highest median nPVIV of the three speakers assessed as multiethnolectal in formal speech is Malik at 47.4 (underlined). The highest mean nPVIV of the three speakers assessed as multiethnolectal in formal speech is Hayder at 56.0 (underlined).

I therefore have highlighted in gray all median nPVIV values less than 47.4 and all mean nPVIV values less than 56.0. The speakers who have two gray boxes are also the speakers who I personally assess as sounding multiethnolectal: Reman, Malik, Max, Antonio, Hayder, Dawit, Murad, Abel, Reza, Mezdar, Jocke, Shorty, and Tarik. The speakers who have one gray box or none also correspond with speakers who, in my ears, sound more mainstream Stockholmian.

The point of this exercise is to test the link between the thin information I have on perceptions of multiethnolect to these findings on production and see whether any patterns emerge. Here, I think a crude pattern is available, but it is far from definitive. It shows an approximate link between working-class and lower middle-class affiliations, *invandrare* identity, assessments of multiethnolect, and low nPVIV. The next step would be to build off of these and construct a perception study that manipulates rhythm in a series of matched-guise stimuli.

6 Focusing and the role of generation 2.5

In contact scenarios, dialect establishment has traditionally been ascribed to the third generation of speakers (Roberts 2000; Singler 2006; Trudgill 2004). The third generation accelerates focusing because the inputs it receives during language acquisition are from the newly-coined systematic dialect of generation 2. Generation 2, on the other hand, is limited in how much variation it can eliminate because its inputs are both the parents' heritage language and highly irregular L2 inputs.

Newer work, however, has drawn attention to the role of younger members of the second generation engaging in the type of work typically ascribed to generation 3. Sharma and Sankaran (2011) offer evidence of this in their analysis of two cohorts within the same second generation of British Asians in Southall, UK. The younger of the second generation produced overall higher rates of British Asian features than the older generation. They point toward changes in the social ecology as the reasons behind this generational split.

> For the older group, surviving at school and in public involved an ability to downplay Indianness and to pass as British. This could be achieved both by acquiring nativelike use of British variants and possibly by beginning to "weaken" Asian variants phonetically, for instance, by favoring postalveolar variants. [...] The younger group had far less regular direct contact with India, and by the time they were growing up, there was a local British Asian community to ori-

> ent to. Fewer direct ties account for their lack of fluent bidialectal ability in Indian English, and the lack of physical threat, hostility, or even substantial contact with non-Asians, reduced the need to pass as "purely" British as well. Both factors encouraged the use of occasional emblematic markers of Punjabiness integrated into British English, clearly distinct in form and function from what they see as "freshie" Indian English. For them, their primary affiliation is their local British Asian peer group, and a focused, Punjabi-inflected British English speech style suits this target well (Sharma & Sankaran 2011:423-424).

Similarly, my own findings show a difference between two groups from the same generation. As a reminder, the results were that the older *invandrare* group produces significantly lower rhythmic alternation than *svensk* speakers and that this is attributed to the reduction of accented vowels. The younger *invandrare* group produces even lower rhythmic alternation than the older *invandrare* group and *svensk* speakers. The effect is implemented by means of partially restoring the size of accented vowels while also enlarging unstressed vowels by means of duration, f_0 and intensity.

I will refer to this young group as 'generation 2.5' due to the fact that this cohort has had different life experiences from generation 2. Two social phenomena are responsible for this fact. First, recalling my earlier discussion of migration waves on page 240, a large spike of refugees entered Sweden 24 years after the first spike of post-war labor migrants. This did not just happen in Sweden. A similar spike can be seen in Denmark and Norway on page 22. A similar spike also occurred in the Germany (Green 2013) and the European Union as a whole (Coppel, Dumont, & Visco 2001:8). This spike contributed to a wave of new Swedish-born children who would have had three key inputs during their Critical Period of language acquisition – the early version of Rinkeby Swedish, the highly variable Swedish interlanguage of all the new arrivals in their community, and their parents' heritage languages.

Second, the gap between generations in the late-modern Global North is wider than it has ever been in human history. For this trend, Scandinavia is no exception. Corne (1995) and Roberts (2000), in their analyses of creole formations, have defined a generation by 17 to 18 years. In contrast, Statistics Sweden reports that the average age of a first mother in 2000 was 28.2 (the earliest I could find data on). For fathers, it was 30.7. What this means is that this new wave of migrants and their Swedish-born children would have had a numerical advantage over the handful of third-generation children being born to the adult children of the first wave.

As a heuristic exercise, let us hypothetically reverse the scenario. Let us envision an average generation size of 17 years instead of 30. By 1994, when the second migration spike occurred, this newly arrived feature pool would have been in stiff competition with the focused features of a large critical mass of third-generation children. Not only would these children have constituted a larger share of the population, they would have also had stronger vernacular inputs from parents who themselves are closer in proximity to youth culture. In fact, the narrow window between generations would result in generally denser ties throughout the community between speakers of different ages. This would likely have rendered the focused variety more immune to new inputs from any second migration surge.

What my results (and Sharma and Sankaran's 2011 results) imply is that in late-modern contact scenarios, we can no longer assume that the third generation will always be as strong in relevance. This is because the gaps are larger between generations, and the new arrivals much more frequent. This does not necessarily mean that the dialects will not stabilize. If one recalls Kerswill and Williams's (1994) findings on /t/ in Milton Keynes, the rates of non-standard variants are remarkably different between 12-year-olds and 4-year-olds and reflect rapid norming (reproduced in Labov 2001:426). In Stockholm's new suburbs, one could expect similar rates of norming, given the similarities between them and Milton Keynes in terms of having no founder population. Nonetheless, an open question is whether focusing will occur or whether the steady stream of new inputs will result in continued divergence.

In either case, we would benefit from finding a way to conceptually partition a seemingly endless second generation, especially since the incremental overlap of micro-cohorts seems more pivotal than the turnover of full cohorts. This is why I suggest the term 'generation 2.5'[3].

7 Rinkeby Swedish 2.0: A lovechild of bourgeois politics?

The findings in Chapter 9 show that 'staccato' is produced more dramatically by younger speakers than older speakers. It reflects a shift from reduced prominent vowels to enlarged unstressed vowels and constitutes an intense divergence from mainstream Swedish. Typologically, it moves the vernacular of Stockholm's racialized working class even further away from every other variety in Stockholm.

The age break among speakers coincides with several important transformations in Swedish society. These are, specifically, an increase in migration, an increase in the income divide, the deregulation of housing and schooling, and the subsequent segregation of neighborhoods and schools in Stockholm. Yet, the establishment that pushed to increase migration[4], restructure the tax code, liberalize Sweden, and restore traditional bourgeois institutions is the same establishment that often is least accepting of the resulting linguistic creation. The macrosocial changes discussed in Chapter 9 were part of a sweeping series of political overhauls by the *Alliance* (Alliansen), a political coalition formed of The Centre Party (Centerpartiet), The

[3] 2.5, however, is also inadequate if we consider the fact that the 'second generation' has been further extended by the recent spike in migration from Syria.

[4] In years 1991–1993, for example, The Christian Democrats and The Liberal People's Party were more receptive to refugees than The Left and The Social Democrats (Demker & Gilljam 1993:229), the latter of which were wary of migration's effect on collective bargaining. The Moderates were only slightly less receptive than the Social Democrats (Demker & Gilljam 1993:229).

Liberal People's Party (Folkpartiet), The Christian Democrats (Kristdemokraterna), and The Moderate Party (Moderaterna). Enelo (2012) found that support for these parties correlated heavily with high economic and symbolic capital (2012:177). Professions that were found to be central within the cluster of high economic and symbolic capital were attorneys, doctors, senior managers, middle managers, and small business owners (2012:148). These constitute the core of the bourgeois establishment that, for the first time in 70 years, managed to reshape public policy according to their more liberalized value system.

Bijvoet (2018) found that this same group of high-capital professionals is the least receptive to the new varieties born out of Sweden's reshaped social landscape. In her study of gatekeepers' receptiveness to Stockholm's social dialects, their receptiveness to multiethnolect in casual speech was more negative than their receptiveness to traditional working-class Swedish and learner Swedish. Some comments that senior managers came up with included the following:

> 20. Astrid: here come my judgements (.) immigrant lad with a sore ego (.) these are guys who think they are great (.) that they are going to make it big even if they haven't done well in school (.) and if you question their abilities they get angry [about Ibrahim]
>
> 21. Alva: he works as a telephone or electricity salesman (.) nice that we aren't stereotyping [about Serre]
>
> 22. Alice: a little bit mañana mentality [laugh] (.) takes each day as it comes, and why work at all [laugh] [about Serre][i] (my translation, Bijvoet 2018:167)

While somewhat ironic, it is not a new phenomenon that hegemonic actors show aversion to the inevitable evolutionary spandrels of their hegemony. This aversion is a reflection of what sociologists refer to as the *Janus Face*, a signature characteristic of neoliberalism (Barker 2013; Swyngedouw 2005; Wacquant 2004, 2010). According to Wacquant (2004),

> [the Janus Face] serves opposite functions for the two collectives it binds in a relation of asymmetric dependency. For the dominant category, its rationale is to confine and control, which translates into what Max Weber calls the "exclusionary closure" of the dominated category. For the latter, however, it is an integrative and protective device insofar as it relieves its members from constant contact with the dominant and fosters consociation and community building within the constricted sphere of intercourse it creates. Enforced isolation from the outside leads to the intensification of social exchange and cultural sharing inside. (Wacquant 2004:5)

Wacquant's account sounds a lot like a description of the social and linguistic data collected in this study. In analyzing school (p. 241), neighborhood (p. 168), and social-network demographics (p. 171, p. 170), it is evident that the *invandrare* working class is one of the most socially-enclosed of the groups in the dataset. The evolution of an even more typologically-distant rhythmic pattern is, in my opinion, a direct result of the "intensification of social exchange and cultural sharing inside" that Wacquant is referring to (2004:5).

Recall from Figure 9.4 (p. 240) that Sweden experienced an even greater spike in migration between 2015 and 2018, mostly due to migrants from Syria. What will be exciting to see is whether this will coalesce with Stockholm's even starker segregation to produce a similar tipping effect within the speech of the racialized precariat. Will there be a Rinkeby Swedish 3.0 or does the current vernacular now fill so many socio-symbolic gaps that incoming substrates will simply be blocked? My hope is that linguists are on the ground now collecting adolescent speech data so that we can answer these important questions in the future.

Citations in original language

[i] 20. Astrid: här kommer fördomen (.) invandrar kille med ömt ego (.) det är killar som tycker att de är bra (.) att de kommer att lyckas även om de inte har klarat skolan (.) och när man ifrågasätter deras möjligheter blir de arga [om Ibrahim]

21. Alva: han jobbar som mobilabonnemang- eller elavtalsförsäljare [skratt] (.) skönt att vi inte kategoriserar [om Serre]

22. Alice: lite mañana-mentalitet [skratt] (.) hantar dagen som den kommer och vad ska man jobba som då inget alls [skratt] [om Serre] (Bijvoet 2018:167)

Bibliography

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition*, *34*(2), 187–214.

Adams, C. (1979). *English speech rhythm and the foreign learner, Janua Linguarum, Series Practica 69, Studia Memoriae Nicolai van Wijk Dedicata*. The Hague: Mouton Publishers.

Adli, A. (2013). Syntactic variation in French Wh-questions: A quantitative study from the angle of Bourdieu's sociocultural theory. *Linguistics*, *51*, 473–515.

Aktürk-Drake, M. (2018). Hur bra har den svenska integrationskontexten varit på att främja balanserad tvåspråkighet? [How well has the Swedish integration context been for advancing balanced bilingualism?]. *Nordand - Nordic Journal of Bilingualism Research*, *2*(02), 107–130.

Allen, G. D. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, *3*, 75–86.

Allwood, J. (1999). The Swedish Spoken Language Corpus at Göteborg University. In *Proceedings Fonetik 99: The Swedish Phonetics Conference, June 1999*. Gothenburg: Gothenburg University.

Andersson, K. (2015, March 12). Özz Nûjen om kritikerna: 'Fuck them all!' [Özz Nûjen about his critics: 'Fuck them all!'. *SVT Online*. Retrieved from http://www.svt.se/kultur/husblatte

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1–2), 46–63.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373.

Asu, E. L., Nolan, F., & Schötz, S. (2017). Rhythm in Estonian Swedish. In J. K. Jardar Eggesbö Abrahamsen & W. van Dommelen (Eds.), *Nordic Prosody, Proceedings of the XIIth Conference, Trondheim 2016* (pp. 73–82). Bern, Switzerland: Peter Lang. Retrieved from https://www.peterlang.com/view/9783631723548/chapter-006.xhtml

Auer, P., & Hinskens, F. (2005). The role of interpersonal accommodation in a theory of language change. In P. Auer, F. Hinskens, & P. Kerswill (Eds.), *Dialect change: Convergence and divergence in European languages* (pp. 335–357). Cambridge: Cambridge University Press.

Austin, J. L. (1962). *How to do things with words Oxford University Press*. Oxford: Oxford University Press.

Baayen, R. H. (1996). The effects of lexical specialization on the growth curve of the vocabulary. *Computational Linguistics*, *22*(4), 455–480.

Baayen, R. H., Endresen, A., Janda, L. A., Makarova, A., & Nesset, T. (2013). Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian linguistics*, *37*(3), 253–291.

Back, L., Dikec, M., Gilroy, P., Listerborn, C., Molina, I., Sernhede, O., … Vradis, A. (2013, June 1). Husby-kommission krävs för djup analys [Husby commisson demanded for deep analysis]. *Svenska Dagbladet*. Retrieved from http://www.svd.se/husby-kommission-kravs-for-djup-analys/om/debatt

Bailey, B. (2000). Language and negotiation of ethnic/racial identity among Dominican Americans. *Language in Society*, *29*(4), 555–582.

Baksi, M. (1982, August 30). Invandrarnas barn - svagast av de svaga [The immigrants' children - weakest of the weakest]. *Aftonbladet*.

Balasubramanian, T. (1980). Timing in Tamil. *Journal of Phonetics*, *8*(4), 449–467.

Barker, V. (2013). Nordic Exceptionalism revisited: Explaining the paradox of a Janus-faced penal regime. *Theoretical Criminology*, *17*(1), 5–25.

Bauman, Z. (1992). *Intimations of Postmodernity*. London: Routledge.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204.

Benor, S. B. (2010). Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity. *Journal of Sociolinguistics*, *14*(2), 159–183.

Bergman, G. (1946). Språket på Söder [The Language on Södermalm]. In G. Bergman (Ed.), *Ord och Stil* [Words and Style] (pp. 100–119). Stockholm: Prisma.

Bergs, A. (2005). *Social Networks and Historical Sociolinguistics: Studies in Morphosyntactic Variation in the Paston Letters (1421–1503)* [Topics in English Linguistics 51]. Berlin & New York: Walter de Gruyter.

Bertinetto, P. M. (1989). Reflections on the dichotomy 'stress' vs.'syllable-timing'. *Revue de phonétique appliquée*, *91–92-93*, 99–130.

Bijvoet, E. (2018). Förortssvenska i grindvakters öron: Perceptioner av migrationsrelaterad språklig variation bortom inlärarspråk och förortsslang [suburban Swedish in the ears of gatekeepers: Perceptions of migration-related linguistic variation beyond interlanguage and suburban slang]. *Språk och stil*, *28*, 142–175.

Bijvoet, E., & Fraurud, K. (2008). Svenskan i dagens flerspråkiga storstadsmiljöer: En explorativ studie av unga stockholmares perceptioner av variation och varieteter [Swedish in today's multilingual enviroments: An explorative study of young Stockholmers' perceptions of variation and varieties]. *Nordand - Nordic Journal of Bilingualism Research*, *3*(2), 7–38.

Bijvoet, E., & Fraurud, K. (2010). Rinkeby Swedish in the mind of the beholder. Studying listener perceptions of language variation in multilingual Stockholm. In P. Quist & B. A. Svendsen (Eds.), *Multilingual Urban Scandinavia - New Linguistic Practices* (pp. 170–188). Toronto: Multilingual Matters.

Bijvoet, E., & Fraurud, K. (2011). Language variation and varieties in contemporary multilingual Stockholm: An exploratory pilot study of young people's perceptions. In R. Källström & I. Lindberg (Eds.), *Young Urban Swedish. Variation and change in multilingual settings. Göteborgsstudier i nordisk språkvetenskap, 14* [Gothenburg studies in Nordic Linguistics, 14] (pp. 1–34). Gothenburg: Gothenburg University Press.

Bijvoet, E., & Fraurud, K. (2012). Studying high-level (L1–L2) development and use among young people in multilingual Stockholm. *Studies in Second Language Acquisition*, *34*(2), 291–319.

Bijvoet, E., & Fraurud, K. (2013). Rinkebysvenska och andra konstruktioner av språklig variation i dagens flerspråkiga Sverige [Rinkeby Swedish and other constructions of linguistic variation in today's multilingual Sweden]. In H. Keneth & L. I (Eds.), *Svenska som andraspråk i forskning, undervisning och samhälle, 2* [Swedish as a second language in research, instruction and society, 2] (pp. 369–396). Lund: Studentlitteratur.

Bijvoet, E., & Fraurud, K. (2016). What's the target? A folk linguistic study of young Stockholmers' constructions of linguistic norm and variation. *Language Awareness*, *25*(1–2), 17–39.

Bjørnskov, C. (2008). Social trust and fractionalization: A possible reinterpretation. *European Sociological Review*, *24*(3), 271–283.

Björsten, S., & Engstrand, O. (1999). Swedish 'damped' /iː/ and /yː/: experimental and typological observations. In *Proceedings of the XiVth International Congress of Phonetic Sciences* (Vol. 3, pp. 1957–1960).

Blom, J.-P., & Gumperz, J. J. (1972 [1986]). Social meaning in Linguistic Structure: Code-Switching in Norway. In J. J. Gumperz & D. Hymes (Eds.), *Directions in Sociolinguistics* (pp. 407–434). Oxford: Basil Blackwell.

Bloomfield, L. (1933). *Language*. London: George Allen & Unwin Ltd.

Bodén, P. (2004, December). A new variety of Swedish. In *Proceedings of the Tenth Australian International Conference on Speech Science and Technology* (pp. 475–480). Sydney: Macquarie University.

Bodén, P. (2005). The sound of 'Swedish on Multilingual Ground'. In *Proceedings of FoNeTiK 2005* (pp. 37–40).

Bodén, P. (2007). 'Rosengårdssvensk' fonetik och fonologi [Rosengård Swedish' phonetics and phonology]. In L. Ekberg (Ed.), *Språket hos ungdomar i en flerspråkig miljö i Malmö* [Youth language in a multilingual milieu in Malmö], Nordlund 27 (pp. 1–

47). Lund: Lund University Press.

Bodén, P. (2010). Pronunciation in Swedish multietnolect. In P. Quist & B. A. Svendsen (Eds.), *Multilingual Urban Scandinavia - New Linguistic Practices* (pp. 65–78). Toronto: Multilingual Matters.

Bodén, P. (2011). Adolescents' pronunciation in multilingual Malmö, Gothenburg and Stockholm. In R. Källström & I. Lindberg (Eds.), *Young Urban Swedish. Variation and change in multilingual settings. Göteborgsstudier i nordisk språkvetenskap, 14* [Gothenburg studies in Nordic Linguistics, 14] (pp. 35–48). Gothenburg: Gothenburg University Press.

Boersma, P., & Weenink, D. (2017). *Praat: doing phonetics by computer* [Computer software], Version 6.0.36. Retrieved from http://www.praat.org/

Böhlmark, A., Holmlund, H., & Lindahl, M. (2015). *Skolsegregation och skolval* [School segregation and school choice]. Stockholm: Institute for the Evaluation of Labour Market and Education Policy (IFAU). Retrieved from http://www.ifau.se/Upload/pdf/se/2015/r-2015-05-Skolsegregation-och-skolval.pdf

Borevi, K. (2014). Multiculturalism and welfare state integration: Swedish model path dependency. *Identities*, *21*(6), 708–723.

Bortoni-Ricardo, S. M. (1985). *The urbanization of rural dialect speakers – A sociolinguistic study in Brazil.* Cambridge: Cambridge University Press.

Bourdieu, P. (1980). The Aristocracy of Culture. *Media, Culture & Society*, *2*, 225–254.

Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. Boston: Harvard University Press.

Bourdieu, P. (1987). What makes a social class? On the theoretical and practical existence of groups. *Berkeley Journal of Sociology*, *32*, 1–17.

Bourdieu, P. (1991). *Language and Symbolic Power*. Boston: Harvard University Press.

Boyd, S., & Fraurud, K. (2010). Challenging the homogeneity assumption in language variation analysis: Findings from a study of multilingual urban spaces. In P. Auer & J. E. Schmidt (Eds.), *Language and Space. An International Handbook of Linguistic Variation, Volume 1: Theories and Methods.* (pp. 686–706). Berlin and New York: Mouton de Gruyter.

Bråmå, Å. (2006). 'White flight'? The production and reproduction of immigrant concentration areas in Swedish cities, 1990-2000. *Urban studies*, *43*(7), 1127–1146.

Brato, T. (2015, July 14). *TB-Basic Vowel Analysis, Version 2.2* [Computer software]. Retrieved from http://www.gnu.org/licenses/gpl.txt

Breadmore, H. L., Krott, A., & Olson, A. C. (2014). Agreeing to disagree: Deaf and hearing children's awareness of subject–verb number agreement. *The Quarterly Journal of Experimental Psychology*, *67*(3), 474–498.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32.

Bresin, R., & Battel, G. U. (2000). Articulation Strategies in Expressive Piano Performance Analysis of Legato, Staccato, and Repeated Notes in Performances of the Andante Movement of Mozart's Sonata in G Major (K 545). *Journal of New Music Research*, *29*(3), 211–224.

Britain, D. (1992). Linguistic change in intonation: The use of high rising terminals in New Zealand English. *Language Variation and Change*, *4*(1), 77–104. doi: 10.1017/S0954394500000661

Brown, A. (1988). The staccato effect in the pronunciation of English in Malaysia and Singapore. *New Englishes: The case of Singapore*, 115–128.

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2014). Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics*, *66*, 45–60.

Brown, R., & Gilman, A. (1960). The Pronouns of Power and Solidarity. In T. A. Sebeok (Ed.), *Style in Language* (pp. 253–276). Cambridge: MIT Press.

Bruce, G. (1983). On rhythmic alternation. *Working Papers in Linguistics*, *25*(1), 35–52.

Brücke, E. W. (1871). *Die physiologischen Grundlagen der neuhochdeutschen Verskunst* [The Physiological Bases of the new High German Art of Verse]. Wien: Carl Gerold's Sohn.

Burling, R. (1981). Black English, blandspråk and halvspråkighet. In *Tvåspråkighet: föredrag från tredje Nordiska tvåspråkighetssymposiet juni 1980* [Bilingualism: presentations from the third Nordic bilingualism symposium June 1980] (pp. 4–5). Umeå: Umeå universitet.

Cameron, R. (2000). Language change or changing selves? Direct quotation strategies in the Spanish of San Juan, Puerto Rico. *Diachronica*, *17*(2), 249–292.

Carlsson, G. (1958). *Social mobility and class structure*. Lund: CWK Gleerup.

Cedergren, H. C. J. (1973). *The interplay of social and linguistic factors in Panama* (Unpublished doctoral dissertation). Cornell University, Ithaca, New York, USA.

Cerquiglini, B. (2001). Le français d'aujourd'hui, ça bouge [Today's French is on the move]. *Construire 7*.

Chambers, J. K., & Trudgill, P. (1998). *Dialectology* (2nd ed.). Cambridge: Cambridge University Press.

Chapman, J. (1972 [1818]). The Music or Melody and Rhythmus of Language. In *Omond.* Hildesheim: Georg Olms Verlag.

Cheshire, J. (1982). *Variation in an English dialect: A sociolinguistic study* (Cambridge Studies in Linguistics, 37). Cambridge: Cambridge University Press.

Cheshire, J., Fox, S., Kerswill, P., & Torgersen, E. (2008). Ethnicity, friendship network and social practices as the motor of dialect change: Linguistic innovation in London. *Sociolinguistica*, 1–23.

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics*, *15*(2), 151–196.

Cheshire, J., Nortier, J., & Adger, D. (2015). Emerging multiethnolects in Europe. *Queen Mary's Occasional Papers Advancing Linguistics*, *33*.

Cindy, G., Ogay, T., & Giles, H. (2005). Communication accommodation theory. In W. B. Gudykunst (Ed.), *Theorizing about Intercultural Communication* (pp. 121–148). Los Angeles: Sage.

Classe, A. (1939). *The Rhythm of English Prose*. Oxford: Blackwell.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, *74*(368), 829–836.

Clyne, M. (2000). Lingua franca and ethnolects in Europe and beyond. *Sociolinguistica*, *14*, 83–89.

Cofer, T. (1972). *Linguistic variability in a Philadelphia speech community* (Unpublished doctoral dissertation). University of Pennsylvania.

Coggshall, E. L. (2008). The prosodic rhythm of two varieties of Native American English. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 1–10.

Coppel, J., Dumont, J.-C., & Visco, I. (2001). *Trends in immigration and economic consequences*. Paris: OECD Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd.org/migration/46608122.pdf

Corne, C. (1995). Nana k nana, nana k napa: The paratactic and hypotactic relative clauses of Reunion Creole. *Journal of Pidgin and Creole Languages*, *10*(1), 57–76.

Cornips, L. (2019, forthcoming). Dialect acquisition by 'new speakers' of Dutch and their linguistic othering. *International Journal of the Sociology of Language*.

Cornips, L., & de Rooij, V. A. (2013). Selfing and othering through categories of race, place, and language among minority youths in Rotterdam, The Netherlands. In S. P., I. Gogolin, M. E. Schulz, & J. Davydova (Eds.), *Multilingualism and language diversity in urban areas: Acquisition, identities, space, education 1* (pp. 129–164). Amsterdam: John Benjamins.

Crompton, A. (1980). Timing patterns in French. *Phonetica*, *37*(4), 205–234.

Crowder, K. (2000). The racial context of white mobility: An individual-level assessment of the white flight hypothesis. *Social Science Research*, *29*(2), 223–257.

Crul, M., Keskiner, E., & Lelie, F. (2017). The upcoming new elite among children of immigrants: a cross-country and cross-sector comparison. *Ethnic and Racial Studies*(2), 209–229.

Cumming, R. E. (2010). *Speech rhythm: The language-specific integration of pitch and duration* (Unpublished doctoral dissertation). University of Cambridge, Cambridge, United Kingdom.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of memory and language*, *25*(4), 385–400.

Danmarks Radio 3. (2006). *Perkerdansk* [Paki Danish]. Copenhagen. Retrieved from https://www.youtube.com/watch?v=oJE9KHsMiYk

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62.

de Beauvoir, S. (2009 [1949]). *The Second Sex* (C. Borde & S. Malovany-Chevallier, Trans.). New York: Vintage Books.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. *Language and Language Processing: Proceedings of the 38th Linguistic Colloquium, Peter Lang*, 231–241.

Dellwo, V. (2010). *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence* (Unpublished doctoral dissertation). Rheinischen Friedrich-Wilhelms-Universität, Bonn, Germany.

Demker, M., & Gilljam, M. (1993). Om rädslan för det främmande [On fear of the foreign]. In M. Demker (Ed.), *Svensk migrationspolitisk opinion 1991—2012* [Swedish migration political opinion 1991–2012] (Vol. 11, pp. 225–236). Gothenburg: SOM-institutet.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, *29*(2), 217–230.

di Paolo, M. (1988). Pronunciation and categorization in sound change. In K. Ferrara, B. Brown, K. Walters, & J. Baugh (Eds.), *Linguistic Change and Contact: Proceedings from NWAV-XVI* (pp. 84–92). Austin: Department of Linguistics, University of Texas.

Dodsworth, R., & Benton, R. A. (2017). Social network cohesion and the retreat from Southern vowels in Raleigh. *Language in Society*, *46*(3), 371–405.

Donaldson, T. (1992). Kant's global rationalism. In T. Nardin & D. R. Mapel (Eds.), *Traditions of International Ethics* (pp. 136–57). Cambridge: Cambridge University Press.

Drager, K., & Kirtley, M. J. (2016). Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 1–24). Cambridge: Cambridge University Press.

Duckworth, J. (1967). The Rhythm of Spoken English. In D. C. Wigglesworth (Ed.), *Selected Conference Papers of The Association of Teachers of English as a Second Language, December, 1967*. Los Altos, California: Language Research Associates' Press.

Duez, D., & Casanova, M.-H. (2000). Quelques aspects de l'organisation temporelle du parlé des banlieues parisiennes. *Revue Parole*(1), 59–73.

Eckert, P. (1989a). *Jocks and burnouts: Social categories and identity in the high school*. New York: Teachers College Press.

Eckert, P. (1989b). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, *1*(3), 245–267.

Eckert, P. (2000). *Language variation as social practice: The linguistic construction of identity in Belten High*. Oxford: Wiley-Blackwell.

Eckert, P. (2008). Where do ethnolects stop? *International Journal of Bilingualism*, *12*(1–2), 25–42.

Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, *41*, 87–100.

Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual review of Anthropology*, *21*(1), 461–488.

Eisikovits, E. (1981). *Inner-Sydney English: An investigation of grammatical variation in adolescent speech* (Unpublished doctoral dissertation). University of Sydney.

Ekberg, L. (2007). "... sån svensk å blond å sånt du vet." Lexiko-grammatiska drag i Malmöungdomars talspråk ["... sån svensk å blond å sånt du vet." Lexico-grammatical features in Malmö youths' spoken language]. In L. Ekberg (Ed.), *Språket hos ungdomar i en flerspråkig miljö i Malmö. Nordlund 27* [The language of youth in a multilingual environment in Malmö]. (pp. 48–77). Lund: Lund University Press.

Ekberg, L. (2011). Extended uses of sån ('such') among adolescents in multilingual Malmoö, Sweden. In R. Källström & I. Lindberg (Eds.), *Young Urban Swedish. Variation and change in multilingual settings. Göteborgsstudier i nordisk språkvetenskap, 14* [Gothenburg studies in Nordic Linguistics, 14] (pp. 49–66). Gothenburg: Gothenburg University Press.

Ekenstam, C., Frykman, J., Johansson, T., Kuosmanen, J., Ljunggren, J., & Nilsson, A. (1998). *Rädd att falla: Studier i Manlighet* [Afraid to fall: Studies in Masculinity]. Hedemora: Gidlund.

Eklund, I., & Traunmüller, H. (1997). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica*, *54*(1), 1–21.

Eliaso Magnusson, J., & Stroud, C. (2012). High proficiency in markets of performance. *Studies in second language acquisition*, *34*(2), 321–345.

Enelo, J.-M. (2012). *Klass, åsikt och partisympati : det svenska konsumtionsfältet för politiska åsikter* (No. 15). Örebro universitet.

Engstrand, O. (1990). Swedish. *Journal of the International Phonetic Association*, *20*(1), 42–44.

Engstrand, O. (1995). Phonetic interpretation of the word accent contrast in Swedish. *Phonetica*, *52*(3), 171–179.

Engstrand, O. (1997). Phonetic interpretation of the word accent contrast in Swedish: Evidence from spontaneous speech. *Phonetica*, *54*(2), 61–75.

Engstrand, O., Bannert, R., Bruce, G., Elert, C.-c., & Eriksson, A. (1998-2003). *SweDia: Phonetics and phonology of Swedish dialects around the year 2000*. Lund, Umeå and Stockholm: Lund University, Umeå University, Stockholm University.

Engstrand, O., & Krull, D. (1994). Durational correlates of quantity in Swedish, Finnish and Estonian: Cross-language evidence for a theory of adaptive dispersion. *Phonetica*, *51*(1–3), 80–91.

Erikson, R., & Goldthorpe, J. H. (1992). *The constant flux: A study of class mobility in industrial societies*. Oxford: Oxford University Press.

Eriksson, A. (1991). *Aspects of Swedish speech rhythm* (Unpublished doctoral dissertation). University of Gothenburg, Gothenburg, Sweden.

ESS. (2012). *European Social Survey, Round 6 Data.* Data file edition 2.1. Norwegian Social Science Data Services, Norway - Data Archive and distributor of ESS data for ESS ERIC.

Eurostat. (2016). *Urban Europe - statistics on cities, towns and suburbs - foreign-born persons living in cities.* Luxembourg: Eurostat. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/Urban_Europe_%E2%80%94_statistics_on_cities,_towns_and_suburbs_%E2%80%94_foreign-born_persons_living_in_cities#Foreign-born_persons_living_in_cities

Experian Ltd, & InsightOne Nordic AB. (2013). *Mosaic Sweden E-Handbook: The Classification of Swedish Consumers.*

Ezilon. (2009). *Geografisk karta över Sverige.* Retrieved from https://sv.maps-sweden.com/sverige-geografi-karta (Accessed: 2019-08-19)

Fagyal, Z. (2010). Rhythm types and the speech of working-class youth in a banlieue of Paris: The role of vowel elision and devoicing. In D. R. Preston & N. Niedzielski (Eds.), *A Reader in Sociophonetics* (pp. 91–132). Berlin: Mouton de Gruyter.

Fagyal, Z., & Stewart, C. (2011). Prosodic style-shifting in preadolescent peer-group interactions in a working-class suburb of Paris. In F. Kern & M. Selting (Eds.), *Ethnic Styles of Speaking in European Metropolitan Areas* (pp. 75–99). Amsterdam, Philadelphia: John Benjamins.

Fagyal, Z., Swarup, S., Escobar, A. M., Gasser, L., & Lakkaraju, K. (2010). Centers and peripheries: Network roles in language change. *Lingua*, *120*(8), 2061–2079.

Fant, G., & Kruckenberg, A. (1989). Preliminaries to the study of Swedish prose reading and reading style. *Dept. for Speech, Music and Hearing Quarterly Progress and Status Report (STL-QPSR)*, *30*(2), 1–80.

Fant, G., & Kruckenberg, A. (1994). Notes on stress and word accent in Swedish. *Dept. for Speech, Music and Hearing, Quarterly Progress and Status Report*, 125–144.

Fant, G., Kruckenberg, A., & Liljencrants, J. (2000). Acoustic-phonetic Analysis of Prominence in Swedish. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 55–86). Dordrecht, Netherlands: Springer.

Farnetani, E., & Kori, S. (1983). Interaction of syntactic structure and rhythmical constraints on the realization of word prosody. *Quaderni del Centro di Studio per le Ricerche di Fonetica*, *2*, 288–318.

Fasold, R. W. (1969). *A sociolinguistic study of the pronunciation of three vowels in Detroit speech*. Washington, D.C.: Center for Applied Linguistics, Sociolinguistics Program.

Feagin, C. (1979). *Variation and change in Alabama English: A sociolinguistic study of the white community*. Washington D.C.: Georgetown University Press.

Fischer, J. L. (1958). Social influences on the choice of a linguistic variant. *Word*, *14*(1), 47–56.

Fontanella de Weinberg, M. B. (1974). *Un aspecto sociolingüístico del español bonaerense: La -s en Bahía Blanca* [A sociolinguistic feature in Bonaerense Spanish: The -s in Bahia Blanca]. Cuadernos de linguistica: Universidad Nacional del Sur.

Fores. (2015). *Sverige* [Sweden]. Retrieved from www.migrationsinfo.se/migration/sverige/

Forsberg, H. (2018). School competition and social stratification in the deregulated upper secondary school market in Stockholm. *British Journal of Sociology of Education*, *39*(6), 891–907.

Forsberg, I. (2014, December 4). Sverige olikt de nordiska länderna i synen på invandring [Sweden unlike the other Nordic countries in its view on immigration]. *Sveriges Radio, P6*.

Fraisse, P. (1963). *The Psychology of Time*. New York: Harper & Row.

Fraisse, P. (1982). Rhythm and Tempo. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 149–180). New York: Academic Press.

Fraurud, K. (2003). Svenskan i Rinkeby och andra flerspråkiga bostadsområden [Swedish in Rinkeby and other multilingual residential environments]. *Sprog i Norden* [Language in Scandinavia], *34*(1), 77–92.

Fraurud, K., & Boyd, S. (2011). The native-non-native speaker distinction and the diversity of linguistic profiles of young people in multilingual urban contexts in Sweden. In R. Källström & I. Lindberg (Eds.), *Young Urban Swedish. Variation and change in multilingual settings. Göteborgsstudier i nordisk språkvetenskap, 14* [Gothenburg studies in Nordic Linguistics, 14] (pp. 67–88). Gothenburg: Gothenburg University Press.

Fromont, R., & Hay, J. (2012). LaBB-CAT: An annotation store. In *Proceedings of the Australasian Language Technology Association Workshop 2012* (pp. 113–117).

Fry, D. B. (1955). Duration and Intensity as Physical Correlates of Linguistic Stress. *The Journal of the Acoustical Society of America*, *27*(4), 765–768.

Fry, D. B. (1958). Experiments in the Perception of Stress. *Language and Speech*, *1*(2), 126–152.

Fuchs, R. (2013). *Speech rhythm in educated Indian English and British English* (Unpublished doctoral dissertation). Westfälische Wilhelms-Universität, Münster, Germany.

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. London: Hamish Hamilton.

Fulton, L. (2013). *Worker representation in Europe*. Labour Research Department and ETUI. SEEurope Network. Retrieved from http://www.worker-participation.eu/National-Industrial-Relations

Gabrielsson, A. (1982). The Performance of Music. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 501–602). New York: Academic Press.

Gal, S. (1979). *Language Shift: Social Determinants of Linguistic Change in Bilingual Austria*. New York & London: Academic Press.

Ganuza, N. (2008). *Syntactic Variation in the Swedish of Adolescents in Multilingual Urban Settings – Subject-verb Order in Declaratives, Questions and Subordinate Clauses* (Vol. 15). Stockholm: Elanders.

Ganuza, N. (2011). Syntactic variation in the Swedish of adolescents in multilingual urban settings – a thesis summary. In R. Källström & I. Lindberg (Eds.), *Young Urban Swedish. Variation and change in multilingual settings. Göteborgsstudier i nordisk språkvetenskap, 14* [Gothenburg studies in Nordic Linguistics, 14] (pp. 89–104). Gothenburg: Gothenburg University Press.

Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A Standard International Socio-Economic Index of Occupational Status. *Social Science Research*, *21*(1), 1–56.

Ganzeboom, H. B., & Treiman, D. J. (2003). Three internationally standardised measures for comparative research on occupational status. In J. H. Hoffmeyer-Zlotnik, J. H. Hoffmeyer-Zlotnik, & C. Wolf (Eds.), *Advances in cross-national comparison: A European working book for demographic and socio-economic variables* (pp. 159–193). New York: Kluwer Academic/Plenum Publishers.

Ganzeboom, H. B., & Treiman, D. J. (2018). *International Stratification and Mobility File: Conversion Tools.* Amsterdam: Department of Social Research. (Accessed: 2018-05-31 from http://www.harryganzeboom.nl/ismf/index.htm)

Gates, S. M. (2018). Why the Long FACE?: Ethnic Stratification and Variation in the London Diphthong System. *University of Pennsylvania Working Papers in Linguistics*, *24*(2), 38–48. Retrieved from https://repository.upenn.edu/pwpl/vol24/iss2/6/

Gibbon, D. (2003). *Computational modelling of rhythm as alternation, iteration and hierarchy.* Barcelona.

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. In D. Aalborg (Ed.), *Proceedings of Eurospeech 2001* (pp. 91–94).

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, 87–105.

Giles, H., & Phillip, S. (1979). Accommodation theory: Optimal levels of convergence. In H. Giles & R. St. Clair (Eds.), *Language and Social Psychology* (pp. 45–65). Oxford: Basil Blackwell.

Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, *39*(3), 192–193.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology*, *7*, 515–546.

Grammont, M. (1932). *Traité Pratique de Prononciation Française* [Treaty on French Pronunciation Practice]. Paris: Librairie Delagrave.

Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, *1*, 201–233. Retrieved from www.jstor.org/stable/202051

Green, S. (2013). Germany: A changing country of immigration. *German Politics*, *22*(3), 333–351.

Gregersen, F., & Pedersen, I. L. (1991). *The Copenhagen study in urban sociolinguistics, Parts I and II.* Copenhagen: C. A. Reitzels Forlag.

Grömping, U. (2009). Variable Importance Assessment in Regression: Linear Regression versus Random forest. *The American Statistician*, *63*(4), 308–319.

Gross, J., Boyd, S., Leinonen, T., & Walker, J. A. (2016). A tale of two cities (and one vowel): Sociolinguistic variation in Swedish. *Language Variation and Change*, *28*(2), 225–247.

Gross, J., & Forsberg, J. (2017). *När två blir ett? Ett begynnande sammanfall av /i/ och /y/ i göteborgska* [When two becomes one? An incipient merger of /i/ and /y/ in Gothenburg Swedish].

Gustafsson, B., & Jansson, B. (2008). Top incomes in Sweden over three-quarters of a century: a micro data approach. *Cambridge Journal of Economics*, *32*(6), 963–976.

Gustavsson, M., & Jordahl, H. (2008). Inequality and trust in Sweden: Some inequalities are more harmful than others. *Journal of Public Economics*, *92*(1–2), 348–365.

Gut, U. (2005). Nigerian English prosody. *English World-Wide*, *26*(2), 153–177.

Haeri, N. (1996). *The Sociolinguistic Market of Cairo: Gender, Class and Education*. London: Kegan Paul.

Hall, T., & Vidén, S. (2005). The Million Homes Programme: A review of the great Swedish planning project. *Planning Perspectives*, *20*(3), 301–328.

Hall-Lew, L. (2010). *Ethnicity and phonetic variation in a San Francisco neighborhood* (Unpublished doctoral dissertation). Department of Linguistics, Stanford University.

Hansegård, N. E. (1968). *Tvåspråkighet eller halvspråkighet?* [Bilingualism or semilingualism?]. Stockholm: Bonniers förlag.

Hansen, G. F., & Pharao, N. (2010). Prosody in the Copenhagen multiethnolect. In P. Quist & B. A. Svendsen (Eds.), *Multilingual Urban Scandinavia - New Linguistic Practices* (pp. 79–95). Toronto: Multilingual Matters.

Harvey, D. (2007). *A Brief History of Neoliberalism*. Oxford University Press, USA.

Hauser, R. M., & Warren, J. R. (1997). Socioeconomic Indexes for Occupations: A Review, Update, and Critique. *Sociological Methodology*, *27*(1), 177–298.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Headey, B. W. (1978). *Housing policy in the developed economy: The United Kingdom, Sweden and the United States*. London: Taylor & Francis.

Hedin, K., Clark, E., Lundholm, E., & Malmberg, G. (2012). Neoliberalization of housing in Sweden: Gentrification, filtering, and social polarization. *Annals of the Association of American Geographers*, *102*(2), 443–463.

Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics*, *31*(1), 39–62.

Heldner, M., & Strangert, E. (1997). To what extent is perceived focus determined by Focues? In *Proceedings of Eurospeech 1997* (pp. 875–877). Rhodes, Greece: ESCA.

Heldner, M., & Strangert, E. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics*, *29*(3), 329–361.

Helgason, P., Ringen, C., & Suomi, K. (2013). Swedish quantity: Central Standard Swedish and Fenno-Swedish. *Journal of Phonetics*, *41*(6), 534–545.

Hellberg, M. (Producer). (2006). Stockholmska [Stockholmian] [Television series episode].

In *Svenska dialektmysterier* [Swedish dialect mysteries]. Stockholm: Sveriges Television. Retrieved from https://www.youtube.com/watch?v=PWRkWFOkiSg

Hellberg, M. (Producer). (2012). Rikssvenska [Realm Swedish] [Television series episode]. In *Svenska dialektmysterier* [Swedish dialect mysteries]. Stockholm: Sveriges Television. Retrieved from https://www.youtube.com/watch?v=mlMhJH7pJsc

Heo, M., & Leon, A. C. (2010). Sample sizes required to detect two-way and three-way interactions involving slope differences in mixed-effects linear models. *Journal of Biopharmaceutical Statistics*, *20*(4), 787–802.

Hesse, B. (2007). Racialized modernity: An analytics of white mythologies. *Ethnic and Racial Studies*, *30*(4), 643–663.

Hewitt, R. (1986). *White talk, black talk: Inter-racial friendship and communication amongst adolescents*. Cambridge: Cambridge University Press.

Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change*, *22*(1), 37–67.

Hollingshead, A. B., & Redlich, F. C. (1958). *Social class and mental illness: Community study*. Hoboken: John Wiley & Sons Inc.

Holmlund, H., Häggblom, J., Lindahl, E., Martinson, S., Sjögren, A., Vikman, U., & Öckert, B. (2014). *Decentralisering, skolval och fristående skolor: Resultat och likvärdighet i svensk skola* [Decentralization, school choice, and charter schools: Results and equivalency in the Swedish school]. Uppsala: Institutet för arbetsmarknads-och utbildningspolitisk utvärdering [Institute for Evaluation of Labour Market and Education Policy] (IFAU).

Hong, Y. (1991). *A sociolinguistic study of Seoul Korean*. Seoul: Research Center for Peace and Unification of Korea.

Hotelling, H. (1933). *Analysis of a Complex of Statistical Variables into Principal Components.* Baltimore: Warwick & York.

Hout, M., Smith, T. W., & Marsden, P. V. (2015). *Prestige and Socioeconomic Scores for the 2010 Census Codes*. Chicago: National Opinion Research Centre (NORC).

Hozo, S. P., Djulbegovic, B., & Hozo, I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*, *5*(1), 13.

Hübinette, T., Hörnfeldt, H., Farahani, F., & Rosales, R. L. (2012). Om ras och vithet i det samtida Sverige [On race and whiteness in contemporary Sweden]. In T. Hübinette, H. Hörnfeldt, F. Farahani, & R. L. Rosales (Eds.), *Om ras och vithet i det samtida Sverige* [On race and whiteness in contemporary Sweden] (pp. 11–36). Stockholm: Mångkulturellt Centrum.

Hyltenstam, K. (1992). Non-native Features of Near-native Speakers: On the Ultimate Attainment of Childhood L2 Learners. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals* (Vol. 83, pp. 351–368). North-Holland. Retrieved from http://

www.sciencedirect.com/science/article/pii/S0166411508615058 doi: https://doi.org/10.1016/S0166-4115(08)61505-8

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, *23*(2), 225–257.

Ingvar, S., & Jellart, O. (1959). *Klokt och tokigt, korta stycken för skolans tal-och läsövningar* [Smart and crazy - short pieces for school speech and reading exercises] (3rd ed.). Stockholm: Carlssons.

International Labour Office. (2012). *International Standard Classification of Occupations 2008 (iScO-08), Volume 1: Structure, group definitions and correspondence tables.* Geneva: International Labour Office.

Jahangiri, N. (1980). *A sociolinguistic study of Tehrani Persian* (Unpublished doctoral dissertation). University of London.

Jakobson, R., Fant, C. G. M., & Halle, M. (1951). *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates.* Cambridge, MA: The MIT Press. Retrieved from http://www.speech.kth.se/gunnarfant/Jakobson_Fant_Halle_Preliminaries_to_Speech_Analysis.pdf

Jerkert, B. (1981, September 26). Barnen talar egna språk - som ingen annan förstår [The children speak their own language - that no one else understands]. *Dagens Nyheter*.

Johnstone, B., & Kiesling, S. F. (2008). Indexicality and experience: Exploring the meanings of/aw/-monophthongization in Pittsburgh. *Journal of Sociolinguistics*, *12*(1), 5–33.

Jonsson, R. (2015). *Värst i klassen: Berättelser om stökiga pojkar i innerstad och förort* [Worst in the class: Narratives of unruly boys in the center city and suburb]. Stockholm: Ordfront.

Källström, R., & Lindberg, I. (Eds.). (2011). *Young Urban Swedish. Variation and change in multilingual settings. Göteborgsstudier i nordisk språkvetenskap, 14* [Gothenburg studies in Nordic linguistics, 14]. Gothenburg: Gothenburg University Press.

Kann, V. (2010). KTHs morfologiska och lexikografiska verktyg och resurser. *LexicoNordica*(17).

Kann, V., & Hollman, J. (2011). *Slutrapport för projektet Vidareutveckling av Folkets lexikon*.

Karlgren, B. (1915). *Études sur la phonologie chinoise* [Studies on the phonology of Chinese]. Stockholm: Norstedt & Söner.

Keim, I. (2007). *Die "türkischen Powergirls". Lebenswelt und kommunikativer Stil einer Migrantinnengruppe in Mannheim* ["The Turkish Powergirls". The lived experiences and communicative style of immigrant girls in Mannheim]. Tübingen: Narr.

Kern, F. (2013). *Rhythmus und Kontrast im Türkischdeutschen* [Rhythm and contrast in Turkish German]. Berlin: Walter de Gruyter.

Kerswill, P., & Williams, A. (2002). 'Salience' as an explanatory factor in language change: evidence from dialect levelling in urban England. In M. C. Jones & E. Esch (Eds.), *Language Change. The interplay of internal, external and extra-linguistic factors* (pp. 81–110). Berlin: De Gruyter Mouton.

Kerswill, P. E., & Williams, A. (1994). *A New Dialect in a New City: Children's and Adults'*

Speech in Milton Keynes. Final report to Economic and Social Research Council.

Kerswill, P. E., & Williams, A. (2000). Creating a new town koine: Children and language change in Milton Keynes. *Language in Society*, *29*(1), 65–115.

Kirkham, S. (2013). *Ethnicity, social practice and phonetic variation in a Sheffield secondary school* (Unpublished doctoral dissertation). School of English Literature, Language and Linguistics, University of Sheffield.

Kirkham, S., & Moore, E. (2016). Constructing social meaning in political discourse: Phonetic variation and verb processes in Ed Miliband's speeches. *Language in Society*, *45*(1), 87–111.

Kisler, T., Reichel, U., Schiel, F., Draxler, C., Jackl, B., & Pörner, N. (2016, may). BAS Speech Science Web Services - an Update of Current Developments. In N. C. C. Chair) et al. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris, France: European Language Resources Association (ELRA).

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, *59*(5), 1208–1221.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038–1054.

Kolinsky, R., Morais, J., & Cluytens, M. (1995). Intermediate Representations in Spoken Word Recognition; Evidence from Word Illusions. *Journal of Memory and Language*, *34*(1), 19–40.

Kotsinas, U.-B. (1988a). Rinkebysvenska - en dialekt? [Rinkeby Swedish - a dialect?]. In P. Linell, V. Adelswärd, T. Nilsson, & P. Petersson (Eds.), *Svenskans beskrivning 16* [The description of Swedish 16] (Vol. 1, pp. 264–278). Linköping: Tema Kommunikation.

Kotsinas, U.-B. (1988b). Stockholmspråk i förändring. In G. Pettersson (Ed.), *Studier i svensk språkhistoria* (Vol. 1). Lund: Lund University Press.

Kotsinas, U.-B. (1990). Svensk, Invandrarsvensk eller Invandrare? Om Bedömning av "Främmande" Drag i Ungdomsspråk [Swedish, Immigrant Swedish or Immigrant? On Assessments of "Foreign" Features in Youth Language]. In G. Tingbjörn (Ed.), *Andra symposiet om svenska som andraspråk i Göteborg 1989* [Second Symposium on Swedish as a Second Language in Gothenburg 1989] (pp. 244–275). Stockholm: Scriptor.

Kotsinas, U.-B. (1994a). Snobbar och pyjamastyper: Ungdomskultur, ungdomsspråk och gruppidentiteter i Stockholm [Snobs and Pyjama types: Youth culture, youth language and group identities in Stockholm]. In J. Fornäs, U. Boethius, M. Forsman, H. Ganetz, & B. Reimer (Eds.), *Ungdomskultur i Sverige* [Youth Culture in Sweden] (pp. 311–336). Stockholm: Brutus Östlings Bokförlag.

Kotsinas, U.-B. (1994b). *Ungdomspråk (Ord och stil 25)* [Youth language (Words and style 25)]. Stockholm: Hallgren & Fallgren.

Kotsinas, U.-B. (2000). Pidginization, creolization and creoloids in Stockholm, Sweden. In B.-L. Gunnarsson, S. Hellberg, & K. Svartholm (Eds.), *Kontakt, variation och förändring - studier i Stockholmsspråk. Ett urval av uppsatser av Ulla-Britt Kotsinas* [Contact, variation and change - studies in the language of Stockholm. A selection of papers of Ulla-Britt Kotsinas] (pp. 160–190). Stockholm: Almqvist & Wiksell International.

Kotsinas, U.-B. (2001). *Lever Ekenssnacket? Om äldre och nyare Stockholmsslang* [Does Ekenssnack live? On older and newer Stockholm slang]. Uppsala: Kungl. Gustav Adolfs Akademien för svensk folkkultur.

Krull, D. (2012). Rhythmic Variability and Swedish-Estonian Language Contact. In *Proceedings of FONETIK 2012* (pp. 81–84).

Kuronen, M. (2000). *Vokaluttalets akustik i sverigesvenska, finlandssvenska och finska (Studia Philologica Jyväskyläensia 49).* Jyväskylä, Finland: University of Jyväskylä Press.

Kuronen, M., & Leinonen, K. (2001). *Fonetiska skillnader mellan finlandssvenska och rikssvenska* [Phonetic differences between Finland Swedish and Realm Swedish]. Linköping.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*(3), 273–309.

Labov, W. (1966a). Hypercorrection by the lower middle class as a factor in linguistic change. In W. Bright (Ed.), *Sociolinguistics* (pp. 105–130). The Hague: Mouton.

Labov, W. (1966b). *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Labov, W. (1972a). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1972b). Some principles of linguistic methodology. *Language in Society*, *1*(1), 97–120.

Labov, W. (1973). The linguistic consequences of being a lame. *Language in Society*, *2*(1), 81–115.

Labov, W. (1993). *The unobservability of structure and its linguistic consequences.*

Labov, W. (1994). *Principles of linguistic change, Volume 1: Internal factors*. Oxford: Blackwell.

Labov, W. (2001). *Principles of linguistic change, Volume 2: Social factors*. Oxford: Blackwell.

Labov, W. (2006 [1966]). *The Social Stratification of English in New York City* (2nd ed.). Cambridge: Cambridge University Press.

Labov, W. (2007). Transmission and Diffusion. *Language*, *83*(2), 344–387.

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., & Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, *15*(4), 431–463.

Ladefoged, P. (1967). Stress and respiratory activity. In P. Ladefoged (Ed.), *Three Areas of Experimental Phonetics* (pp. 46–47). London: Oxford University Press.

Lauzen, M. M., & Dozier, D. M. (2005). Maintaining the double standard: Portrayals of age and gender in popular films. *Sex Roles*, *52*(7–8), 437–446.

Lavandera, B. (1975). *Linguistic structure and sociolinguistic conditioning in the use of verbal endings in 'SI' clauses (Buenos Aires, Argentina)* (Unpublished doctoral dissertation).

University of Pennsylvania.

Lebrun, Y. (1966). Is Stress Essentially a Thoracic or an Abdominal Pulse? A Finding of 'Not Proven'. In Y. Lebrun (Ed.), *Linguistic Research in Belgium*. Wetteren.

Lee, R. L. (2010). On the margins of belonging: Confronting cosmopolitanism in the late modern age. *Journal of Sociology*, *46*(2), 169–186.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Massachusetts: M.I.T. Press.

Lehiste, I. (1972). The Timing of Utterances and Linguistic Boundaries. *The Journal of the Acoustical Society of America*, *51*(6:2), 2018–2024.

Leinonen, T. (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects*. Groningen: Groningen Dissertations.

Lemberg, U. (1978, November 17). Bilden av Sverige. födda till diskare, städare och spärrvakter [The picture of Sweden. Born to become dishwashers, janitors and token-booth clerks]. *Aftonbladet*.

Lentin, A. (2008). Europe and the Silence about Race. *European journal of social theory*, *11*(4), 487–503.

Lentin, A., & Titley, G. (2011). *The crises of multiculturalism: Racism in a neoliberal age*. London, New York: Zed Books.

Lerdahl, E., & Jackendoff, R. A. (1983). *Generative theory of tonal music*. Cambridge, Massachusetts: MIT Press.

Levelt, W. J. (1989). *Speaking: From intention to articulation*. Cambridge, Massachusetts: M.I.T. Press.

Levon, E., & Fox, S. (2014). Social Salience and the Sociolinguistic Monitor: A Case Study of ING and TH-fronting in Britain. *Journal of English Linguistics*, *42*(3), 185–217.

Li Wei. (1994). *Three generations, two languages, one family: Language choice and language shift in a Chinese community in Britain*. Philadelphia: Multilingual Matters.

Lieberman, M. D. (2003). Reflexive and reflective judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. D. Williams, & W. von Hippel (Eds.), *Social Judgments: Implicit and Explicit Processes* (pp. 1–34). Cambridge: Cambridge University Press.

Lieberman, P. (1960). Some Acoustic Correlates of Word Stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451–454.

Lindblom, B. (1975). Some Temporal Regularities of Spoken Swedish. In G. Fant & M. A. A. Tatham (Eds.), *Auditory Analysis and Perception of Speech* (pp. 387–396). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/{B}9780122485503500295

Lindblom, B., & Rapp, K. (1971). Re-examining the compensatory adjustment of vowel duration in Swedish words. *Speech Transmission Laboratory, Quarterly Progress and Report*, *12*(4), 19–25.

Lobanov, B. M. (1971). Classification of Russian Vowels Spoken by Different Speakers. *The Journal of the Acoustical Society of America*, *49*(2B), 606–608.

Low, E. L. (1998). *Prosodic prominence in Singapore English* (Unpublished doctoral dissertation). University of Cambridge, Cambridge, United Kingdom.

Low, L. E., Grabe, E., & Nolan, F. (2000). Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English. *Language and Speech*, *43*(4), 377–401.

Lundström, C. (2010). Transnationell vithet - Svenska migrantkvinnor i USA och Singapore [Transnational whiteness - Swedish migrant women in USA and Singapore]. *Tidskrift för genusvetenskap*(1–2), 24–45.

Macaulay, R. K. (1976). Social class and language in Glasgow. *Language in Society*, *5*(2), 173–188.

Madsen, E. L., Daumerie, B., & Hardee, K. (2010). The Effects of Age Structure on Development. *Policy and Issue Brief, Population Action International*, 1–4. Retrieved from https://www.issuelab.org/resources/5007/5007.pdf

Manrique, A. M. B., & Signorini, A. (1983). Segmental duration and rhythm in Argentine Spanish. *Journal of Phonetics*, *11*, 117–128.

Matory, J. L. (2015). Stureplan People. *Transition: An International Review*(118), 47–60.

Matsumoto, K. (2010). The role of social networks in the post-colonial multilingual island of Palau: Mechanisms of language maintenance and shift. *Multilingua - Journal of Cross-Cultural and Interlanguage Communication,*, *29*, 133–165.

Mattys, S. L., & Melhorn, J. F. (2005). How do syllables contribute to the perception of spoken English? Insight from the migration paradigm. *Language and Speech*, *48*(2), 223–252.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Proceedings of Interspeech* (pp. 498–502).

Mehrabani, M., Mishra, T., & Conkie, A. (2013). Unsupervised prominence prediction for speech synthesis. *INTERSPEECH-2013*, 1559–1563.

Meyerhoff, M. (2015). *Introducing Sociolinguistics*. London, New York: Routledge.

Meyerhoff, M., & Schleef, E. (2012). Variation, contact and social indexicality in the acquisition of (ing) by teenage migrants 1. *Journal of Sociolinguistics*, *16*(3), 398–416.

Milani, T. M., & Jonsson, R. (2012). Who's Afraid of Rinkeby Swedish? Stylization, Complicity, Resistance. *Journal of Linguistic Anthropology*, *22*(1), 44–63.

Milardo, R. M. (1988). Families and social networks: An overview of theory and methodology. In R. M. Milardo (Ed.), *Families and social networks* (pp. 13–47). Newbury Park: Sage.

Milroy, J., & Milroy, L. (1978). Belfast: Change and variation in an urban vernacular. In P. Trudgill (Ed.), *Sociolinguistic patterns in British English* (pp. 19–36). London: Edward Arnold.

Milroy, L. (1987). *Language and Social Networks* (2nd ed.). Oxford: Basil Blackwell.

Mishra, T., Sridhar, V. R., & Conkie, A. (2012). Word prominence detection using robust yet simple prosodic features. In *Thirteenth Annual Conference of the International Speech Communication Association* (pp. 1–4).

Modaressi, Y. (1978). *A sociolinguistic investigation of modern Persian* (Unpublished doctoral dissertation). University of Kansas.

Mok, P. P. K., & Lee, S. I. (2008). Korean speech rhythm using rhythmic measures. *Paper presented at the 18th International Congress of Linguists (CIL18), Seoul, Korea.* Retrieved from http://www.cuhk.edu.hk/lin/people/peggy/S

Molina, I., & De los Reyes, P. (2002). Kalla mörkret natt! Kön, klass och ras/etnicitet i det postkoloniala Sverige [Call darkness night! Gender, class, and race/ethnicity in postcolonial Sweden]. In P. De los Reyes, I. Molina, & D. Mulinari (Eds.), *Maktens (o)lika förklädnader. kön, klass och etnicitet i det postkoloniala Sverige* [Power's various disguises. Gender, class, and ethnicity in postcolonial Sweden] (pp. 295–317). Stockholm: Atlas förlag.

Morris, U., & Zetterman, H. (2011). *Från bondgård till cirkus. Konstruktion av en högläsningstext för bedömning av röst-och talfunktion och talandning* [From farm to circus. Constructing a reading passage for assessment of voice and speech faculties] (Master's thesis, Department of Speech Therapy, Karolinska Institutet, Stockholm, Sweden). Retrieved from https://clintec.ki.se/Exam_logopedi/pdf/338.pdf

Mulinari, D., & Neergaard, A. (2004). *Den nya svenska arbetarklassen - Facket och de rasifierade arbetarna.* [The new Swedish working class - Trade unions and racialized workers]. Umeå: Borea.

Mulinari, D., & Neergaard, A. (2014). We are Sweden Democrats because we care for others: Exploring racisms in the Swedish extreme right. *European Journal of Women's Studies*, *21*(1), 43–56.

Myrberg, S. (2010). *The intonational phonology of Stockholm Swedish, Acta Universitatis Stockholmiensis.* Stockholm: Stockholm University Press.

Myrberg, S., & Riad, T. (2015). The prosodic hierarchy of Swedish. *Nordic Journal of Linguistics*, *38*(2), 115–147.

Nakatani, L. H., O'Connor, K. D., & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica*, *38*(1–3), 84–105.

Navarro, T. (1932). *Manual de Pronunciación Español* [Manual of Spanish Pronunciation]. Madrid: Centros de Estudios Históricos.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766.

Nazzi, T., Iakimova, G., Bertoncini, J., Frédonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language*, *54*(3), 283–299.

Neergaard, A. (2017). The Swedish Model in transition: Trade unions and racialised workers. In A. Ålund, C.-U. Schierup, & A. Neergaard (Eds.), *Reimagineering the Nation:*

Essays on twenty-first-century Sweden (pp. 85–117). Frankfurt am Main: Peter Lang.

Nordberg, B. (1975). Contemporary social variation as a stage in a long-term phonological change. In K.-H. Dahlstedt (Ed.), *The Nordic languages and modern linguistics 2.* Stockholm: Almqvist & Wiksell.

Nordenstam, K., & Wallin, I. (2002). *Osynliga flickor – synliga pojkar: Om ungdomar med svenska som andraspråk* [Invisible girls - visible boys: On youth with Swedish as a second language]. Lund: Studentlitteratur.

O'Connor, J. D. (1965). The Perception of Time Intervals. *Progress Report, Department of Phonetics, University College, London*, 11–15.

O'Connor, J. D. (1968). The Duration of the Foot in relation to the number of component sound-segments. *Progress Report, Phonetics Laboratory, University College, London.*

O'Connor, J. D. (1973). *Phonetics.* Harmondsworth, Middlesex: Penguin Books.

Odell, J. (1969 [1806]). An Essay on the Elements, Accents, and Prosody of the English Language, London. In *Anglistica & Americana, 47* (p. 153). New York: Georg Olms.

OECD. (2011). *An Overview of Growing Income Inequalities in OECD Countries: Main Findings.* Retrieved from http://www.oecd.org/els/soc/49499779.pdf

OECD. (2012). *Improving Schools in Sweden: An OECD Perspective.* Retrieved from http://www.oecd.org/edu/school/Improving-{S}chools-in-{S}weden.pdf

Olsen, C. L. (1972). Rhythmical patterns and syllabic features of the Spanish sense group. In *Proceedings of the Seventh International Congress of Phonetic Sciences* (pp. 990–995). The Hague: Mouton.

Open Society Foundations. (2014). *Europe's white working-class communities: Stockholm.* New York: Open Society Foundations.

Öqvist, J. (2010). Riktig stockholmska? Ekenssnacket som varietet och fenomen [Real Stockholmian? Ekenssnack as a variety and phenomenon]. In *Studier i svenska språkets historia, 11 (Acta Academiae Regiae Gustavi Adolphi CxIiI)* (pp. 253–260).

Öqvist, J. (2015, May 27–29). Upper class Swedish. In *Paper presented at the International Conference on Language Variation in Europe (ICLaVE) 8, Leipzig, Germany.*

Ostberg, J. (2007). The linking value of subcultural capital: constructing the Stockholm brat enclave. In B. Cova, R. Kozinets, & A. Shankar (Eds.), *Consumer Tribes* (pp. 93–106). Amsterdam: Butterworth-Heinemann.

Parker, S., Mielke, J., & Pennington, F. (2018). Acoustic and articulatory study of /ɨ/ and /ɯ/ in Bora. In *University of North Carolina Spring Colloquium, April 18.*

Patmore, C. (1857). English Metrical Critics. *The North British Review*, *27*.

Pike, K. L. (1945). *The Intonation of American English.* Ann Arbor, Michigan: The University of Michigan Press.

Platt, J. T., & Weber, H. (1980). *English in Singapore and Malaysia: Status, features, functions.* Kuala Lumpur: Oxford University Press.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561–2569.

Podesva, R. J. (2011). Salience and the social meaning of declarative contours: Three case studies of gay professionals. *Journal of English Linguistics*, *39*(3), 233–264.

Popova, S. (2007). *Överklass* [Upper class]. Falun: ScandBook AB.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., … others (2011). *The Kaldi speech recognition toolkit* (Tech. Rep.). Piscataway: IEEE Signal Processing Society.

Preston, D. R. (2010). Belles Body Just Caught the Fit Gnat: The Perception of Northern Cities Shifted Vowels by Local Speakers. In D. R. Preston & N. Niedzielski (Eds.), *A Reader in Sociophonetics* (pp. 241–252). Berlin, Boston: De Gruyter Mouton.

Prieto, P., del Mar Vanrell, M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, *54*(6), 681–702.

Quist, P. (2000). Ny københavnsk 'multietnolekt' - Om sprogbrug blandt unge i sprogligt og kulturelt heterogene miljøer [New Copenhagen 'multiethnolect' - On language use among youth in linguistically and culturally heterogeneous milieus]. *Danske talesprog*, *1*, 143–211.

Quist, P. (2005). *Stilistiske Praksisser i Storbyens Heterogene Skole. En etnografisk og sociolingvistisk undersægelse af sproglig variation* [Stylistic practices in diverse metropolitan schools. An ethnographic and sociolinguistic investigation of linguistic variation] (Unpublished doctoral dissertation). Copenhagen University, Copenhagen, Denmark.

Quist, P. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism*, *12*(1–2), 43–61.

Quist, P. (2012). Skandinavisk i kontakt med indvandrersprog [Scandinavian in contact with immigrant languages]. *Sprog i Norden*, *43*(1).

Rampton, B. (2006). *Language in late modernity: Interaction in an urban school.* Cambridge: Cambridge University Press.

Ramus, F., Nespor, F., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.

Repp, B. (1998). *Perception and Production of Staccato articulation on the Piano.* Retrieved from http://www.haskins.yale.edu/haskins/STAFF/repp.html

Riad, T. (1998). Towards a Scandinavian accent typology. In W. Kehrein & R. Wiese (Eds.), *Phonology and morphology of the Germanic languages, Linguistische Arbeiten 386* (pp. 77–109). Tübingen: Niemeyer.

Riad, T. (2006). Scandinavian accent typology. In T. Stolz (Ed.), *STUF - Language Typology and Universals, Special issue on Swedish* (Vol. 59, pp. 36–55). Oldenburg: Oldenbourg Wissenschaftsverlag GmbH.

Riad, T. (2009). The morphological status of accent 2 in North Germanic simplex forms. In *Nordic Prosody X, Helsinki, 2008* (pp. 205–216). Frankfurt am Main: Peter Lang.

Riad, T. (2014). *The Phonology of Swedish*. Oxford: Oxford University Press.

Rickford, J. R. (1986). The Need for New Approaches to Social Class Analysis in Sociolinguistics. *Language & Communication*, *6*(3), 215–221.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73–79). London: Edward Arnold.

Roberts, S. J. (2000). Nativization and the genesis of Hawaiian Creole. In J. McWhorter (Ed.), *Language change and language contact in pidgins and creoles* (Vol. 21, pp. 257–300). Amsterdam: John Benjamins.

Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (2011). *FAvE (Forced Alignment and Vowel Extraction) Program Suite* [Computer software]. Retrieved from https://babel.ling.upenn.edu/phonetics/old_website_2015/p2fa/index.html

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, *17*(4), 688–690.

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive psychology*, *68*, 33–58.

Sandoval, S., Berisha, V., Utianski, R. L., Liss, J. M., & Spanias, A. (2013). Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America*, *134*(5), EL477–EL483.

Sarmah, P., Gogoi, D. V., & Wiltshire, C. (2011). Thai English – Rhythm and Vowels. In L. Lim & N. Gisborne (Eds.), *Typology of Asian Englishes* (pp. 75–96). Philadelphia: Benjamins.

Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., … Miles, A. (2013). A new model of social class? Findings from the BBC's Great British Class Survey experiment. *Sociology*, *47*(2), 219–250.

Sharma, D. (2011). Style repertoire and social change in British Asian English. *Journal of Sociolinguistics*, *15*(4), 464–492.

Sharma, D. (2017). Scalar effects of social networks on language variation. *Language Variation and Change*, *29*(3), 393–418.

Sharma, D., & Sankaran, L. (2011). Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change*, *23*(3), 399–428.

Shen, Y., & Peterson, G. G. (1962). *Isochronism in English, Studies in Linguistics*. Buffalo, New York: University of Buffalo.

Shousterman, C. (2015). *Speaking English in Spanish Harlem: Language Change in Puerto Rican English* (Unpublished doctoral dissertation). New York University, New York, USA.

Silva-Corvalán, C. (1981). Extending the sociolinguistic variable to syntax: The case of pleonastic clitics in Spanish. In D. Sankoff & H. Cedergren (Eds.), *Variation Omnibus* (pp. 335–342). Alberta: Linguistic Research.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language &*

Communication, *23*(3), 193–229.

Singler, J. V. (2006). Yes, but not in the Caribbean. *Journal of Pidgin and Creole languages*, *21*(2), 337–358.

Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008).*

Söderström, M., & Uusitalo, R. (2010). School Choice and Segregation: Evidence from an Admission Reform. *The Scandinavian Journal of Economics*, *112*(1), 55–76.

Sontag, S. (1969). A letter from Sweden. *Ramparts*, *8*(1), 23–38.

Squires, L. (2016). Processing grammatical differences: Perceiving versus noticing. In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 80–103). Cambridge: Cambridge University Press.

Stadsmätningsavdelningen i Stockholm. (1971). *SE/SsA/Tryckta kartor/Officiell karta över Stockholmstrakten/Blad 22 Spånga k:a, år 1971.*

Ståhle, C. I. (1975). En undersökning av Stockholmsspråk [An investigation of Stockholm speech]. In C. Ståhle (Ed.), *Stockholmsnamn och Stockholmsspråk* [Stockholm names and Stockholm speech] (pp. 81–89). Stockholm: Norstedts.

Statistics Sweden. (2002). *Personer med utländsk bakgrund, Riktlinjer för redovisning i statistiken* [Statistics on persons with foreign background, Guidelines and recommendations]. Stockholm: SCB-Statistiska Centralbyrån [Statistics Sweden].

Statistics Sweden. (2017). *Sweden's population from 1749 to today* [Sveriges folkmängd från 1749 och fram till idag]. Retrieved from https://www.scb.se/hitta-statistik/artiklar/2017/sveriges-folkmangd-fran-1749-och-fram-till-idag/

Steadman, J. (1938). Affected and Effeminate Words. *American Speech*, *13*(1), 13–18.

Steele, J. (1775 [1969]). An Essay towards establishing the melody and measure of speech. In *English Linguistics, 1500–1800*. Menston, England: The Scholar Press Limited.

Stewart, C. M. (2012). On the Socio-Indexicality of a Parisian French Intonation Contour. *Journal of French Language Studies*, *22*(2), 251–271. doi: 10.1017/S0959269511000457

STOR. (2016, May 19). *Svartskaller* [Pakis][music video]. Hanson, N. (director) & Morberg, M. (producer). Retrieved from https://youtu.be/wZRziALIx9c?t=89

Strangert, E. (1985). *Swedish speech rhythm in a cross-language perspective* (Unpublished doctoral dissertation). Umeå University, Umeå, Sweden.

Strangert, E., & Heldner, M. (1995). The labelling of prominence in Swedish by phonetically experienced transcribers. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)* (Vol. 4, pp. 204–207).

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(307), 1471–2105.

Stroud, C. (2004). Rinkeby Swedish and semilingualism in language ideological debates: A Bourdieuean perspective. *Journal of Sociolinguistics*, *8*(2), 196–214.

Svedgård, L. (1980, November 24). Därför knäcktes 16-årige Petro i invandrar-Sverige [That is why 16-year-old Petro cracked in Immigrant-Sweden]. *Aftonbladet*.

Svendsen, B. A., & Røyneland, U. (2008). Multiethnolectal facts and functions in Oslo, Norway. *International Journal of Bilingualism*, *12*(1–2), 63–83.

Svensson, G. (2007). Funktion och betydelse hos *duvet* i två gymnasistgrupper i Malmö" [funtion and meaning of *duvet* among two high-school groups in Malmö"]. In L. Ekberg (Ed.), *Språket hos ungdomar i en flerspråkig miljö i Malmö, Nordlund 27* [The language of youth in a multilingual environment in Malmö]. (pp. 48–77). Lund: Lund University Press.

Svensson, G. (2011). Discourse particles in Swedish youth talk in multilingual settings in Malmö. In R. Källström & I. Lindberg (Eds.), *Young Urban Swedish. Variation and change in multilingual settings. Göteborgsstudier i nordisk språkvetenskap, 14* [Gothenburg studies in Nordic Linguistics, 14] (pp. 219–236). Gothenburg: Gothenburg University Press.

Sveriges Television. (1979). *Invandrarungdomar: En tidsinställd bomb* [Immigrant youth: A ticking time-bomb]. P1 Studio S in TV1 Lektionskväll, October 24.

Sveriges Television. (1985). *Reportage om Rinkebysvenska* [Report on Rinkeby Swedish].

Sweden. (2015). Sweden and Migration. *Official website of Sweden*. Retrieved from www.sweden.se/migration

Swyngedouw, E. (2005). Governance innovation and the citizen: The Janus face of governance-beyond-the-state. *Urban studies*, *42*(11), 1991–2006.

Tagliamonte, S. A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change*, *24*(2), 135–178.

Tay, M. W. J. (1982). The phonology of educated Singapore English. *English Worldwide*, *3*(2), 135–145.

Thelander, M. (2009). Svenskt standardspråk som begrepp och fenomen [Swedish standard language as a concept and phenomenon]. *Norsk lingvistisk tidsskrift* [Norwegian linguistic journal], *27*(1), 179–198. Retrieved from http://web.novus.no/NLT/NLT_2009-1.pdf

Thibault, P. (1983). *Equivalence et grammaticalisation* (Unpublished doctoral dissertation). Universite de Montreal.

Thomas, E. (2011). *Sociophonetics: An Introduction*. New York: Macmillan International Higher Education.

Thomas, E., & Carter, P. (2006). Prosodic rhythm and African American English. *English World-Wide*, *27*, 331–355.

Thomson, W. (1904). *The Basis of English Rhythm*. Glasgow: Holmes.

Tingsell, S. (2011). Tracking language change. Anaphoric binding in multilingual urban settings in Sweden. In R. Källström & I. Lindberg (Eds.), *Young Urban Swedish. Vari-*

ation and change in multilingual settings. Göteborgsstudier i nordisk språkvetenskap, 14 [Gothenburg studies in Nordic Linguistics, 14] (pp. 237–250). Gothenburg: Gothenburg University Press.

Tongue, R. K. (1974). *The English of Singapore and Malaysia*. Singapore: Eastern Universities Press.

Torgersen, E., & Szakay, A. (2012). An investigation of speech rhythm in London English. *Lingua*, *122*(7), 822–840.

Trudgill, P. (1971). *The Social Differentiation of English in Norwich* (Unpublished doctoral dissertation). University of Edinburgh.

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, *1*(2), 179–195.

Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Trudgill, P. (1986). *Dialects in Contact*. Oxford: Basil Blackwell.

Trudgill, P. (1988). Norwich Revisited: Recent Linguistic Changes in an English Urban Dialect. *English World-Wide*, *9*(1), 33–49.

Trudgill, P. (2004). *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America*, *99*(6), 3782–3790.

Ulfsdotter Eriksson, Y. (2006). *Yrke, Status och Genus - En sociologisk studie om yrken på en segregerad arbetsmarknad* [Occupation, Status and Gender - A sociological study of occupations in a segregated labour market] (Unpublished doctoral dissertation). Gothenburg University, Gothenburg, Sweden.

Uslaner, E. M. (2002). *The Moral Foundations of Trust*. Cambridge: Cambridge University Press.

Vayra, M., Avesani, C., & Fowler, C. (1984). Patterns of temporal compression in spoken Italian. In *Proceedings of the Tenth International Congress of Phonetic Sciences* (pp. 541–546).

Veblen, T. (1931). *The Theory of the Leisure Clan*. New York: Viking.

Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies*, *30*(6), 1024–1054.

Wacquant, L. (2004). What Is a Ghetto? Constructing a Sociological Concept. In N. J. Smelser, P. B. Baltes, et al. (Eds.), *International encyclopedia of the social & behavioral sciences*. London: Pergamon Press.

Wacquant, L. (2010). Crafting the neoliberal state: Workfare, prisonfare, and social insecurity. *Sociological Forum*, *25*(2), 197–220.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7–9), 905–945.

Warner, W. L., Meeker, M., & Eells, K. (1949). *Social class in America: A manual of procedure* *for the measurement of social status.* Oxford: Science Research Associates.

Watson, K., & Clark, L. (2013). How salient is the nurse˜ square merger? *English Language & Linguistics*, *17*(2), 297–323.

Watt, P., Millington, G., & Huq, R. (2014). East London Mobilities: The 'Cockney Diaspora' and the Remaking of the Essex ethnoscape. In P. Watt & P. Smets (Eds.), *Mobilities and Neighbourhood Belonging in Cities and Suburbs* (pp. 121–144). London: Springer.

Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. Lehman & Y. Malkiel (Eds.), *Directions for Historical Linguistics* (pp. 97–195). Austin: University of Texas Press.

Wenner, L. (2010). *När lögnare blir lugnare: en sociofonetisk studie av sammanfallet mellan kort ö och kort ui uppländskan* [When lögnare becomes lugnare: a sociophonetic study of the collapse between short ö and short u in Upplandish] (Unpublished doctoral dissertation). Department of Scandinavian Languages, Uppsala University.

Werbart, B. (2000). Vad etnicitet *inte* är [What ethnicity is *not*]. *Fornvännen*, *95*(3), 183–188.

Wesselhoff, J., & Gauding, J. (2015). *Validering av texten "Trapetskonstnären": analys av högläsning av personer med neuromotorisk talstörning och friska kontrollpersoner* [Validation of the text "The trapeze artist": Analysis of reading of persons with neuromotor speech disorders and healthy control speakers] (Master's thesis, Department of Speech Therapy, Karolinska Institutet, Stockholm, Sweden). Retrieved from https://openarchive.ki.se/xmlui/handle/10616/44955

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501–522.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679.

Wiese, H. (2009). Grammatical innovation in multiethnic urban Europe: New linguistic practices among adolescents. *Lingua*, *119*(5), 782–806.

Wiese, H. (2012). *Kiezdeutsch: Ein neuer Dialekt entsteht* [Hood German: A new dialect emerges]. Munich: C.H. Beck.

Wimmer, A. (2008). The making and unmaking of ethnic boundaries: A multilevel process theory. *American Journal of Sociology*, *113*(4), 970–1022.

Wirén, A. (1991, January 22). Helena, 7 år – klassens enda svensk [Helena, 7 years – the only Swede in her class]. *Expressen*.

Wolf, C., & Jiménez, E. (1979). *A sound change in progress: Devoicing of Buenos Aires /ž/.* Ms.

Wolfram, W. (1969). *A Sociolinguistic Description of Detroit Negro Speech, Urban Language Series, No. 5.* Arlington: Center for Applied Linguistics.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology.* New York: Wiley.

Woods, H. B. (1979). *A socio-dialectology survey of the English spoken in Ottawa: A study of sociological and stylistic variation in Canadian English* (Unpublished doctoral dissertation). University of British Columbia.

Woolard, K. A. (1985). Language variation and cultural hegemony: Toward an integration of sociolinguistic and social theory. *American Ethnologist*, *12*(4), 738–748.

Yaeger-Dror, M. (1993). Linguistic analysis of dialect "correction" and its interaction with cognitive salience. *Language Variation and Change*, *5*(2), 189–224.

Young, N. (2014). *Suburban Swedish Maturing* (Unpublished master's thesis). Department of Linguistics, Stockholm University, Stockholm, Sweden.

Young, N. (2018a). 'Copycats, ja dom shouf': Using hip hop to compare lexical replications in Danish and Swedish multiethnolects. *University of Pennsylvania Working Papers in Linguistics*, *24*(2), 174–184.

Young, N. (2018b, Feb 9-10). Introducing NordaS: Automatic Segmentation of Nordic Languages. *presented at Fonologi i Norden (FINO)* [Phonology in the Nordic Countries], Lund, Sweden.

Young, N. (2018c). Talrytmens sociala betydelse i det senmoderna Stockholm [The social meaning of speech rhythm in late-modern Stockholm]. *Nordand - Nordic Journal of Bilingualism Research*, *13*(1), 41–63. Retrieved from https://docs.wixstatic.com/ugd/5215a0_b6223326a63645c4b35811dc4b1501c4.pdf

Young, N., & McGarrah, M. (2017, November 2-5). Introducing NordfA - Forced Alignment of Nordic Languages. *Presentation at New Ways of Analyzing Variation (NWAV) 46, Madison, USA*.

Young, S., Woodland, P., & Byrne, W. (1993). *HTK: Hidden Markov Model Toolkit V1. 5*. Cambridge: Cambridge Univ. Eng. Dept. Speech Group and Entropic Research Lab. Inc.

Ystén, H. (2008). Stockholm har blivit kallt [Stockholm has become cold]. *Offside*, *6*, 70–86. Retrieved from www.offside.org

Yuan, J., Ryant, N., Liberman, M., Stolcke, A., Mitra, V., & Wang, W. (2013). Automatic phonetic segmentation using boundary models. In F. Bimbot et al. (Eds.), *INTERSPEECH 2013, Proceedings of the 14th Annual Conference of the International Speech Communication Association, Lyon, France, August 25-29* (pp. 2306–2310). Retrieved from http://www.isca-speech.org/archive/interspeech_2013

Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, *37*(2), 231–247.

Appendix A

Information sheet and consent form

Queen Mary
University of London

Queen Mary Ethics of
Research Committee
Ref: QMERC2015/72

Informationsbrev

Forskningsprojekt: *Vardagen, bekantskaper och det svenska språket i Stockholm*

Hej!

Tack för ditt intresse för mitt forskningsprojekt. Mitt namn är Nate Young, och den studie jag vill att du deltar i är en del av mitt avhandlingsprojekt i sociolingvistik (samhällsinriktad språkvetenskap) på Queen Mary, University of London.

Syfte, innehåll och tidsinsats för din medverkan

Syftet med projektet är att öka förståelsen av stockholmsspråket. Det finns en allmän brist på ny forskning om det svenska språket, så dina insikter har stor betydelse för detta forskningsområde. Jag vill gärna säkerställa att ditt samtycke är 100% frivilligt och informerat, och därför vill jag att du läser nedanstående information noga.

Basmedverkan består av två sittningar:

1. Ett inspelat gruppsamtal (ungefär 1 timme) där gruppen får en rad ämneskort att snacka om, vilket syftar till att väcka ens mest avslappnade pratsätt. Därför är ämnena också luftiga som, till exempel, <<*Är Stockholm en wannabe världstad?*>> eller <<*Är svenska tjejer snyggare än engelska tjejer?*>>
2. En intervju med mig (ungefär 1 timme) om din sociala bakgrund (t.ex., utbildning, uppväxt, mm.) och dina språkkunskaper (t.ex., hur många språk du talar och hur bra, mm.). Det finns även några korta högläsningstexter och frågor om din musiksmak och klädstil.

- (Sen om du är intresserad, finns det en tredje sittning om dina sociala kontakter, vilket tar mellan 45 minuter och 1 timme.)

Mötesplatsen bestämmer vi ihop beroende på dina preferenser. Det enda viktiga är att jag kan spela in oss utan bakgrundsbrus (ljudinspelning med diktafon). Man kompenseras med 100 kr finansierat av The *British Economic and Social Research Council* (http://www.esrc.ac.uk/). Man får också en kopia av den färdiga avhandlingen.

Analys och samhällsimplikationer

Studien flätar ihop kvalitativa och kvantitativa metoder, och därför är din engagerade medverkan på djupet viktig för mig. För att kunna förstå hur stockholmsspråket ser ut idag, måste man erkänna individens komplexitet: en individ ändrar sitt pratsätt beroende på situation, och variationen kan bero på ens uppväxt, kontakter och även ens personliga identitetspolitik. För att kunna identifiera de faktorer som påverkar språkvariation mest, behöver studien få en grundlig bild av varje deltagare. Med mer än 30 deltagare och graden av "djup-data" hoppas jag kunna svara på forskningsfrågor som tidigare studier inte lyckats med. Om du är nyfiken berättar jag gärna mer om min forskning när vi ses.

Tystnadsplikt

Jag behandlar uppgifterna från studien med strikt sekretess. Jag har lagstadgad tystnadsplikt beträffande inspelningarna och dina uppgifter. Vid ytterligare frågor, känn dig fri att kontakta mig på +44 7490 212 957 eller på nathan.j.young@qmul.ac.uk. Tack på förhand för din medverkan!

Med varma hälsningar,

Nate Young
doktorand,
Queen Mary, University of London

If you have any questions or concerns about the manner in which the study was conducted, please in the first instance contact the researcher responsible for the study. If this is unsuccessful or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W117, Queen's Building, Mile End Campus, Mile End Road, London, United Kingdom, or research-ethics@qmul.ac.uk.

Queen Mary
University of London

Queen Mary Ethics of
Research Committee
Ref: QMERC2015/72

Samtycke

Forskningsprojekt:

Vardagen, bekantskaper och det svenska språket i Stockholm

Se till att du läst informationsbrevet och att forskaren förklarat projektet till dig.

Namn: ______________________________ Tel. Nr.: ______________________

Mejl: ____

Jag ger mitt samtycke till att delta i följande forskningsprojekt: ***Vardagen, bekantskaper och det svenska språket i Stockholm,*** Queen Mary Ethics of Research Committee Ref: QMERC2015/72, projektledaren Nathan Joel Young.

grundläggande samtycke 1

Signatur: _____________________________ Datum: _______________________

Forskaren har förklarat projektet till mig, och jag har läst informationsbrevet. <u>Jag gör detta frivilligt och förstår att jag kan avbryta min medverkan när som helst om jag vill.</u>

grundläggande samtycke 2

Signatur: _____________________________ Datum: _______________________

Jag förstår att forskarna ska behandla informationen från studien med strikt sekretess och att de har lagstadgad tystnadsplikt. Ifall att uppgifterna beskrivs i avhandlingen eller i publiceringar, ska de anonymiseras med fiktivt namn, fiktivt område, fiktiv sysselssättning, mm., <u>för att göra igenkänning omöjlig</u>. Dessutom ska informationen lagras efter den brittiska Data Protection Act 1988 samt svensk lag.

grundläggande samtycke 3

Signatur: _____________________________ Datum: _______________________

Forskarens garanti:

Jag, Nate Young, bekräftar att jag noggrant förklarat kraven i denna forskning.

Signatur: _____________________________ Datum: _______________________

Jag ger mitt samtycke att min röst får spelas upp <u>anonymt</u> (med fiktivt namn) enbart på akademiska konferenser och lektioner.

kompletterande samtycke

Signatur: _____________________________ Datum: _______________________

☐ Denna person ska <u>INTE</u> ha sin röst spelad upp på akademiska konferenser och lektioner.

If you have any questions or concerns about the manner in which the study was conducted, please in the first instance contact the researcher responsible for the study. If this is unsuccessful or not appropriate, please contact the Secretary at the Queen Mary Ethics of Research Committee, Room W117, Queen's Building, Mile End Campus, Mile End Road, London, United Kingdom, or research-ethics@qmul.ac.uk.

Appendix B

Praat script for extraction of rhythm data

Adapted from Brato (2015).

```
########## Begin Forms ##########

# This form is for the basic settings applying to all the data
form Basic vowel formant analysis for NORM
   comment Directory of sound files (top) and TextGrids (bottom) with final slash:
   # Never forget the trailing slash for the script to work!
   text sound_directory: /Users/nathanjyoung/Dropbox/DISSERTATION/Analysis/_nPVI/Output/
   text textGrid_directory /Users/nathanjyoung/Dropbox/DISSERTATION/Analysis/_nPVI/Output/

   comment Path to (existing) output file :
   text vowel_outputfile /Users/nathanjyoung/Dropbox/DISSERTATION/Analysis/_nPVI/Output/
        → output_vowels.txt

   comment Overwrite existing output file append data to file?
   boolean Overwrite_output_files 0

   # The vowel labels
   # Remember: Praat is case-sensitive!
   comment Monophthong labels (do not forget the " "):
   text Monophthong_labels " "

   comment Diphthong labels (do not forget the " "):
   text Diphthong_labels " AA0 AA1 AA2 AA3 AA4 AA5 AA6 AA7 AE0 AE1 AE2 AE3 AE4 AE5 AE6 AE7
        → AEE0 AEE1 AEE2 AEE3 AEE4 AEE5 AEE6 AEE7 AEH0 AEH1 AEH2 AEH3 AEH4 AEH5 AEH6 AEH7
        → AH0 AH1 AH2 AH3 AH4 AH5 AH6 AH7 AJ0 AJ1 AJ2 AJ3 AJ4 AJ5 AJ6 AJ7 AU0 AU1 AU2 AU3
        → AU4 AU5 AU6 AU7 EE0 EE1 EE2 EE3 EE4 EE5 EE6 EE7 EH0 EH1 EH2 EH3 EH4 EH5 EH6 EH7
        → EJ0 EJ1 EJ2 EJ3 EJ4 EJ5 EJ6 EJ7 ER0 ER1 ER2 EU0 EU1 EU2 EU3 EU4 EU5 EU6 EU7 IH0
        → IH1 IH2 IH3 IH4 IH5 IH6 IH7 II0 II1 II2 II3 II4 II5 II6 II7 J JJ OA0 OA1 OA2 OA3
        → OA4 OA5 OA6 OA7 OAH0 OAH1 OAH2 OAH3 OAH4 OAH5 OAH6 OAH7 OE0 OE1 OE2 OE3 OE4 OE5
        → OE6 OE7 OEE0 OEE1 OEE2 OEE3 OEE4 OEE5 OEE6 OEE7 OEH0 OEH1 OEH2 OEH3 OEH4 OEH5 OEH6
        →  OEH7 OH0 OH1 OH2 OH3 OH4 OH5 OH6 OH7 OJ0 OJ1 OJ2 OJ3 OJ4 OJ5 OJ6 OJ7 OO0 OO1 OO2
        → OO3 OO4 OO5 OO6 OO7 R RR UH0 UH1 UH2 UH3 UH4 UH5 UH6 UH7 UU0 UU1 UU2 UU3 UU4 UU5
        → UU6 UU7 YH0 YH1 YH2 YH3 YH4 YH5 YH6 YH7 YY0 YY1 YY2 YY3 YY4 YY5 YY6 YY7 "

endform

# Make sure that vowel labels are used correctly
vowel_labels$ = monophthong_labels$ + diphthong_labels$

########## Begin Set up a couple of parameters ##########

# Set up the sound and TextGrid file extensions.
sound_file_extension$ = ".wav"
textGrid_file_extension$ = ".TextGrid"

# Set up the margin (as above)
margin = 0.05

########## End Set up a couple of parameters ##########

########## Begin Load the tables if they exist ##########

# Load the vowel output file or set it up.

if fileReadable (vowel_outputfile$)
   do ("Read Table from tab-separated file...", vowel_outputfile$)

   table_vowel_file = selected ("Table")

   if overwrite_output_files = 1
   beginPause ("Overwrite output files")
   comment ("The vowel output file already exists.")
```

```
   comment ("Are you really sure you want to overwrite it.")
   overwrite_vowel_output = endPause ("Keep file and append new data", "Overwrite", 1)

      if overwrite_vowel_output = 2
         deleteFile (vowel_outputfile$)
         selectObject (table_vowel_file)
         do ("Remove")

         # Create new table
         titleline$ = "Speaker Vowel Context start_time end_time duration mean_dB mean_f0 f1
             → f2 f3 f1g f2g f3g"
         do ("Create Table with column names...", "output_vowels", 0, titleline$)
         table_vowel_file = selected ("Table")

      #endif overwrite_vowel_output = 2
      endif

   # endif overwrite_output_files = 1
   endif

else

   # if the file does not exist, create table file

   titleline$ = "Speaker Vowel Context start_time end_time duration mean_dB mean_f0 f1 f2 f3
       → f1g f2g f3g"
   do ("Create Table with column names...", "output_vowels", 0, titleline$)
   table_vowel_file = selected ("Table")

# endif fileReadable (vowel_outputfile$)
endif

########## End Load the tables if they exist ##########

# Index all sound files in the sound directory set up above
do ("Create Strings as file list...", "list", sound_directory$ + "*" + sound_file_extension$
    → )
string_of_sound_files = selected ("Strings")
number_of_files = do ("Get number of strings")

########## Begin Loop that goes through every file in the input directory ##########

# Note to user that analysis has started.

for file_number to number_of_files

   # Get the filename
   filename$ = do$ ("Get string...", file_number)

   # Open it as a long sound file
   do ("Open long sound file...", sound_directory$ + filename$)
   longsound = selected ("LongSound")

   #Extract the sound name
   speakername$ = replace$ (filename$, sound_file_extension$, "", 0)

   # Open a TextGrid by the same name:
   gridfile$ = "'textGrid_directory$''speakername$''textGrid_file_extension$'"

   if fileReadable (gridfile$)

   # Get maximum formant
   beginPause ("Provide the maximum formant!")
      comment ("Provide the maximum formant for filename: 'speakername$'.")
      comment ("5000 Hz is a typical value for males.")
      comment ("5500-6000 Hz is a typical value for females.")
      comment ("Values for children are usually considerably higher.")
      integer ("Maximum formant", 5000)
   gender_continue = endPause ("Continue", 1)
   if gender_continue = 1

      max_formant = maximum_formant

   endif

   ########## Begin Go through the TextGrid and carry out analyses ##########

   # Open a TextGrid by the same name:
   #gridfile$ = "'textGrid_directory$''speakername$''textGrid_file_extension$'"

   #if fileReadable (gridfile$)
      do ("Read from file...", gridfile$)
      textgrid = selected ("TextGrid")

      # Get the number of tiers
      number_of_tiers = do ("Get number of tiers")

      # Get the name of and number of intervals on tier 1

      number_of_intervals_tier_1 = do ("Get number of intervals...", 1)
```

```
########## Begin Carry out analyses based on tier 1 ##########

# The following loop goes through intervals on tier 1
for interval_tier_1 to number_of_intervals_tier_1

  # Get the label of the tier_1 interval
  tier_1_label$ = do$ ("Get label of interval...", 1, interval_tier_1)

  if index (vowel_labels$, " " + tier_1_label$ + " ")

    # Set up a temporary variable for monophthong or diphthong
    if index (monophthong_labels$, " " + tier_1_label$ + " ")

      monophthong = 1

    else

      monophthong = 0

    #endif index (vowel_labels$, " " + tier_1_label$ + " ")
    endif

    # Set up some variables we'll need below
    segment_mono = 0
    segment_resampled = 0

    ########## Begin Collect the relevant interval data ##########

    ########## Begin Collect tier 1 interval data ##########

    # Get the start, end, duration (in ms) and centre of tier 1 intervals
    tier_1_start = do ("Get starting point...", 1, interval_tier_1)
    tier_1_end = do ("Get end point...", 1, interval_tier_1)
    tier_1_duration = (tier_1_end - tier_1_start)
    tier_1_centre = (tier_1_start + tier_1_end)/2

    # Get the times at various places
    time_twenty = tier_1_start + 20*(tier_1_end-tier_1_start)/100
    time_fifty = tier_1_start + 50*(tier_1_end-tier_1_start)/100
    time_eighty = tier_1_start + 80*(tier_1_end-tier_1_start)/100

    ########## End Collect tier 1 interval data ##########

    ########## Begin Collect tier 2 interval data ##########

    if number_of_tiers > 1

      interval_tier_2 = do ("Get interval at time...", 2, tier_1_centre)

      # Get the label of the tier_2 interval
      tier_2_label$ = do$ ("Get label of interval...", 2, interval_tier_2)

    #endif number_of_tiers > 1
    endif

    ########## End Collect tier 2 interval data ##########

    ########## End Collect the relevant interval data ##########

    ########## Begin Change number of channels and change sampling frequency
        → ##########

    selectObject (longsound)

    # Get the number of channels and sampling frequency
    long_soundinfo$ = Info
    number_of_channels$ = extractLine$(long_soundinfo$, "Number of channels: ")
    number_of_channels = number (number_of_channels$)

    sampling_frequency$ = extractLine$(long_soundinfo$, "Sampling frequency: ")
    sampling_frequency = number (sampling_frequency$)

    # Extract the segment on tier 1 for further analysis
    segment = do ("Extract part...", tier_1_start-margin, tier_1_end+margin, "yes")

    # Check if the segment is stereo and change to mono
    if number_of_channels = 2
      segment_mono = do ("Convert to mono")

    # endif number_of_channels = 2 and convert_to_mono = 1
    endif

    # If the sampling frequency is higher than 16000 Hz, downsample.
    if sampling_frequency > 16000

      current_segment = selected ("Sound")
      segment_resampled = do ("Resample...", 16000, 50)

      # if we resample, we remove the segment with the original sampling rate
      selectObject (current_segment)
```

```
   do ("Remove")
   selectObject (segment_resampled)

# endif sampling_frequency > 16000
endif

########## End Change number of channels and change sampling frequency ##########

########## Begin Collect a range of vowel measurements ##########

   # Get the number of the current sound to select below
   current_segment = selected ("Sound")

   # Get the formants and formant bandwidths
   noprogress do ("To Formant (burg)...", 0.0, 5, max_formant, 0.025, 50)

   ########## Begin Get formant measurements ##########

   # f1 values
   f1_twenty = do ("Get value at time...", 1, time_twenty, "Hertz", "Linear")
   f1_fifty = do ("Get value at time...", 1, time_fifty, "Hertz", "Linear")
   f1_eighty = do ("Get value at time...", 1, time_eighty, "Hertz", "Linear")

   # f2 values
   f2_twenty = do ("Get value at time...", 2, time_twenty, "Hertz", "Linear")
   f2_fifty = do ("Get value at time...", 2, time_fifty, "Hertz", "Linear")
   f2_eighty = do ("Get value at time...", 2, time_eighty, "Hertz", "Linear")

   # f3 values
   f3_twenty = do ("Get value at time...", 3, time_twenty, "Hertz", "Linear")
   f3_fifty = do ("Get value at time...", 3, time_fifty, "Hertz", "Linear")
   f3_eighty = do ("Get value at time...", 3, time_eighty, "Hertz", "Linear")

   ########## Begin Write vowel measurements to tables ##########

   # Get the number of the current sound to select below
   select (current_segment)

   # Get the formants and formant bandwidths
   noprogress do ("To Pitch...", 0, 75, 600)

   mean_f0 = do ("Get mean...", tier_1_start, tier_1_end, "Hertz")

   # Get the number of the current sound to select below
   select (current_segment)

   # Get the formants and formant bandwidths
   noprogress do ("To Intensity...", 80, 0, "yes")

   mean_dB = do ("Get mean...", tier_1_start, tier_1_end, "dB")

   # Remove sound and formant object
   plus current_segment
   do ("Remove")


   selectObject (table_vowel_file)
   do ("Append row")
   last_row_vowels = do ("Get number of rows")

   # Write results to table
   if monophthong = 1

      do ("Set string value...", last_row_vowels, "Speaker", speakername$)
      do ("Set string value...", last_row_vowels, "Vowel", tier_1_label$)
      do ("Set string value...", last_row_vowels, "Context", tier_2_label$)
      do ("Set numeric value...", last_row_vowels, "start_time", 'tier_1_start')
      do ("Set numeric value...", last_row_vowels, "end_time", 'tier_1_end')
      do ("Set numeric value...", last_row_vowels, "duration", 'tier_1_duration')
      do ("Set numeric value...", last_row_vowels, "mean_dB", mean_dB)
      do ("Set numeric value...", last_row_vowels, "mean_f0", mean_f0)
      do ("Set numeric value...", last_row_vowels, "f1", f1_fifty)
      do ("Set numeric value...", last_row_vowels, "f2", f2_fifty)
      do ("Set numeric value...", last_row_vowels, "f3", f3_fifty)
      do ("Set string value...", last_row_vowels, "f1g", " ")
      do ("Set string value...", last_row_vowels, "f2g", " ")
      do ("Set string value...", last_row_vowels, "f3g", " ")

   else

      do ("Set string value...", last_row_vowels, "Speaker", speakername$)
      do ("Set string value...", last_row_vowels, "Vowel", tier_1_label$)
      do ("Set string value...", last_row_vowels, "Context", tier_2_label$)
      do ("Set numeric value...", last_row_vowels, "start_time", 'tier_1_start')
      do ("Set numeric value...", last_row_vowels, "end_time", 'tier_1_end')
      do ("Set numeric value...", last_row_vowels, "duration", 'tier_1_duration')
      do ("Set numeric value...", last_row_vowels, "mean_dB", mean_dB)
      do ("Set numeric value...", last_row_vowels, "mean_f0", mean_f0)
      do ("Set numeric value...", last_row_vowels, "f1", f1_twenty)
      do ("Set numeric value...", last_row_vowels, "f2", f2_twenty)
```

```
            do ("Set numeric value...", last_row_vowels, "f3", f3_twenty)
            do ("Set numeric value...", last_row_vowels, "f1g", f1_eighty)
            do ("Set numeric value...", last_row_vowels, "f2g", f2_eighty)
            do ("Set numeric value...", last_row_vowels, "f3g", f3_eighty)

          #endif monophthong = 1
          endif


        ########## End Write vowel measurements to table ##########

        ########## End Collect a range of vowel measurements ##########

        ########## End Write measurements to tables ##########

      # endif index (vowel_labels$, " " + tier_1_label$ + " ")
      endif

      selectObject (textgrid)

    # endfor interval_tier_1 from 1 to number_of_intervals_tier_1
    endfor

    ########## End Carry out analyses based on tier 1 ##########

    selectObject (textgrid)
    do ("Remove")

  # endif fileReadable (gridfile$)
  endif

  selectObject (longsound)
  do ("Remove")

  # Select the next sound file
  selectObject (string_of_sound_files)

# endfor file_number to number_of_files
endfor

########## End loop that goes through every file in the input directory ##########

# Write the result of the table object to a file

selectObject (table_vowel_file)
do ("Save as tab-separated file...", vowel_outputfile$)
do ("Remove")

selectObject (string_of_sound_files)
do ("Remove")

writeInfo ()
writeInfoLine ("Done!")
appendInfoLine ("")
appendInfoLine ("The results for the vowels analysis have been written to 'newline$''
    → vowel_outputfile$'")
```

Appendix C

BBEdit TextFactory for converting Praat output into Excel

This script merges the measurements for coda /r/ and /j/ with the preceding vowel and was written in BBEdit version 12.5.1 (411060, 64-bit)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE plist PUBLIC "-//Apple//DTD PLIST 1.0//EN" "http://www.apple.com/DTDs/PropertyList
    → -1.0.dtd">
<plist version="1.0">
<dict>
  <key>TextTransformComponents</key>
  <array>
            <dict>
      <key>ComponentArguments</key>
      <dict>
        <key>CaseSensitive</key>
        <true/>
        <key>ReplaceString</key>
        <string>\1\2_\9\3\4\11\t=\11-\4\7\13\14\15</string>
        <key>SearchString</key>
        <string>([\S]+\t)[ ]*([A-Z]+[0-9])[ ]*(\t[\S ]+\t)([\S ]+\t)([\S ]+\t)([\S ]+\t)
            → ([\S\t ]+\n)([\S ]+\t)[ ]*(R|RR|J|JJ)[ ]*(\t[\S ]+\t)\5([\S ]+)(\t[\S\t
            → ]+\n)([\S ]+\t[ ]*[A-Z]+[0-9][ ]*\t[\S ]+\t)(?!\11)([\S\t ]+\n)</string>
        <key>UseGrep</key>
        <true/>
      </dict>
      <key>ComponentName</key>
      <string>ReplaceAll</string>
    </dict>
            <dict>
      <key>ComponentArguments</key>
      <dict>
        <key>CaseSensitive</key>
        <true/>
        <key>ReplaceString</key>
        <string>\1\2_\9\3\4\11\t=\11-\4\7\13\14\15</string>
        <key>SearchString</key>
        <string>([\S]+\t)[ ]*([A-Z]+[0-9])[ ]*(\t[\S ]+\t)([\S ]+\t)([\S ]+\t)([\S ]+\t)
            → ([\S\t ]+\n)([\S ]+\t)[ ]*(R|RR|J|JJ)[ ]*(\t[\S ]+\t)\5([\S ]+)(\t[\S\t
            → ]+\n)([\S ]+\t[ ]*[A-Z]+[0-9][ ]*\t[\S ]+\t)(?!\11)([\S\t ]+\n)</string>
        <key>UseGrep</key>
        <true/>
      </dict>
      <key>ComponentName</key>
      <string>ReplaceAll</string>
    </dict>
    <dict>
      <key>ComponentArguments</key>
      <dict>
        <key>CaseSensitive</key>
        <true/>
        <key>ReplaceString</key>
        <string>\1\2_\9\3\4\11\t=\11-\4\7\13\14\15</string>
        <key>SearchString</key>
        <string>([\S]+\t)[ ]*([A-Z]+[0-9])[ ]*(\t[\S ]+\t)([\S ]+\t)([\S ]+\t)([\S ]+\t)
            → ([\S\t ]+\n)([\S ]+\t)[ ]*(R|RR|J|JJ)[ ]*(\t[\S ]+\t)\5([\S ]+)(\t[\S\t
            → ]+\n)([\S ]+\t[ ]*[A-Z]+[0-9][ ]*\t[\S ]+\t)(?!\11)([\S\t ]+\n)</string>
        <key>UseGrep</key>
        <true/>
      </dict>
      <key>ComponentName</key>
      <string>ReplaceAll</string>
    </dict>
```

```
      <dict>
         <key>ComponentArguments</key>
         <dict>
            <key>CaseSensitive</key>
            <true/>
            <key>ReplaceString</key>
            <string></string>
            <key>SearchString</key>
            <string>([\S ]+\t)(R|RR|J|JJ)(\t[\S ]+\t[\S\t ]+\n)</string>
            <key>UseGrep</key>
            <true/>
         </dict>
         <key>ComponentName</key>
         <string>ReplaceAll</string>
      </dict>
      <dict>
         <key>ComponentArguments</key>
         <dict>
            <key>CaseSensitive</key>
            <true/>
            <key>ReplaceString</key>
            <string></string>
            <key>SearchString</key>
            <string>--undefined--</string>
            <key>UseGrep</key>
            <true/>
         </dict>
         <key>ComponentName</key>
         <string>ReplaceAll</string>
      </dict>
   </array>
   <key>TextTransformOptions</key>
   <dict>
      <key>SaveMode</key>
      <string>ask </string>
      <key>SearchRecursive</key>
      <true/>
      <key>TextFilesOnly</key>
      <true/>
   </dict>
   <key>TextTransformSources</key>
   <array/>
</dict>
</plist>
```

Appendix D

Mosaic market segments

Table of taste assessments for each speaker with *Mosaic* group assigned to speaker at bottom of the table (Experian 2013). Five market segments emerged in the dataset, the descriptions of which are provided after the table.

Interest	Speakers: Abel	August	Antonio	Dawit	Gunnar	Hayder	Johan	Jan-Axel	Jan-Bertel	Jonte	Jocke	Joseph	Jesper	Johnny	Kevin	Malik	Majeed	Martin	Mateo	Max	Mezdar	Murad	Nils	Paul	Per	Parviz	Richie	Rasmus	Reman	Reza	Sven	Sohrab	Solomon	Shorty	Tarik	Thomas
Gardening	Y		Y				Y					Y	Y	Y		Y					Y		Y				Y				Y		Y		Y	
Astrology	Y		Y	Y		Y						Y										Y					Y						Y			
Diet tips																Y	Y		Y										Y							
Visiting sports events	Y	Y		Y		Y			Y	Y	Y					Y	Y	Y	Y	Y		Y	Y	Y	Y	Y		Y	Y	Y			Y	Y	Y	Y
Cinema/Audio equipment	Y	Y	Y	Y					Y			Y					Y								Y				Y					Y	Y	Y
Computers	Y		Y	Y				Y	Y			Y	Y				Y								Y				Y		Y	Y		Y	Y	Y
Union issues				Y						Y		Y			Y		Y	Y				Y								Y					Y	
Football	Y	Y	Y	Y		Y			Y	Y	Y				Y	Y	Y	Y	Y	Y			Y	Y	Y	Y	Y	Y	Y				Y	Y	Y	
Outdoor life	Y	Y	Y	Y		Y	Y		Y	Y	Y	Y	Y			Y		Y			Y	Y	Y			Y	Y		Y	Y	Y	Y	Y	Y	Y	Y
Follow popular TV-shows	Y	Y	Y		Y	Y	Y		Y	Y		Y			Y	Y	Y	Y		Y			Y		Y			Y	Y	Y		Y		Y		Y
Associations		Y		Y	Y			Y			Y				Y	Y	Y	Y			Y		Y							Y			Y	Y	Y	
Golf		Y					Y		Y							Y		Y			Y															
Exercise classes/Gym	Y		Y	Y	Y	Y			Y		Y				Y	Y	Y	Y	Y	Y	Y	Y										Y	Y	Y	Y	Y
Interior decorating	Y		Y				Y		Y			Y	Y	Y			Y				Y	Y					Y		Y				Y	Y	Y	
Health and healthcare	Y		Y	Y	Y	Y			Y		Y				Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y		Y	Y		Y	Y	Y	Y	Y
Healthy Foods	Y		Y	Y	Y	Y			Y		Y			Y	Y		Y	Y	Y	Y	Y		Y	Y	Y				Y	Y		Y	Y	Y	Y	Y
Ice Hockey	Y	Y							Y	Y						Y		Y	Y					Y	Y		Y	Y					Y			Y
Hunting			Y			Y										Y												Y	Y							
Equality	Y	Y		Y	Y		Y		Y	Y		Y	Y		Y		Y	Y	Y			Y	Y		Y	Y				Y	Y	Y	Y	Y	Y	Y
Martial arts	Y		Y		Y	Y					Y		Y		Y	Y			Y	Y	Y	Y		Y				Y	Y	Y		Y	Y	Y	Y	
Art	Y		Y						Y	Y		Y	Y			Y							Y				Y								Y	
Royal families							Y																	Y	Y											
Celebrities							Y																				Y	Y						Y		
Literature	Y									Y		Y	Y			Y		Y					Y				Y								Y	
Skiing							Y	Y								Y		Y					Y	Y												
Reading books	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y		Y	Y				Y	Y	Y		Y		Y	Y		Y	Y	Y	Y	
Crosswords							Y			Y	Y	Y	Y								Y														Y	Y
Sudoku				Y					Y	Y	Y		Y					Y													Y		Y	Y		
Cooking		Y	Y	Y	Y		Y	Y	Y	Y	Y	Y			Y	Y	Y	Y	Y		Y				Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Continued on next page...

...continued from previous page

Interest	Abel	August	Antonio	Dawit	Gunnar	Hayder	Johan	Jan-Axel	Jan-Bertel	Jonte	Jocke	Joseph	Jesper	Johnny	Kevin	Malik	Majeed	Martin	Mateo	Max	Mezdar	Murad	Nils	Paul	Per	Parviz	Richie	Rasmus	Reman	Reza	Sven	Sohrab	Solomon	Shorty	Tarik	Thomas
Meditation/Yoga			Y		Y	Y					Y				Y		Y	Y	Y		Y		Y				Y		Y	Y			Y	Y	Y	
Repairing cars																Y																				Y
Environmental care				Y	Y							Y				Y	Y	Y					Y	Y						Y			Y		Y	Y
Fashion							Y						Y				Y								Y										Y	Y
Exercise	Y	Y	Y	Y	Y	Y		Y	Y		Y				Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y			Y		Y	Y	Y	Y	Y
Motor sport		Y				Y			Y							Y												Y	Y							Y
New technology	Y	Y		Y				Y	Y			Y	Y	Y		Y	Y					Y			Y				Y		Y	Y		Y	Y	Y
National economy	Y	Y		Y	Y		Y	Y	Y			Y			Y	Y	Y	Y			Y	Y		Y		Y				Y		Y	Y	Y	Y	Y
Private economy		Y	Y	Y	Y		Y	Y	Y			Y		Y	Y		Y	Y	Y		Y	Y		Y		Y			Y	Y				Y	Y	Y
Trying new products																																				
Religion			Y	Y	Y	Y	Y			Y		Y	Y		Y	Y		Y	Y		Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Horse riding		Y	Y										Y			Y													Y			Y				
Watch sport on TV	Y	Y	Y	Y		Y			Y	Y	Y				Y	Y	Y	Y	Y	Y			Y		Y			Y		Y		Y	Y	Y	Y	
Sailing			Y		Y							Y				Y									Y		Y									
Holidays in Sweden		Y	Y	Y	Y	Y	Y		Y	Y	Y	Y		Y		Y	Y	Y					Y	Y			Y	Y	Y	Y		Y	Y	Y	Y	Y
Holidays abroad	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y
Disease and health care	Y		Y			Y	Y					Y				Y	Y	Y	Y		Y			Y	Y	Y	Y	Y		Y			Y	Y	Y	Y
Tax				Y	Y		Y	Y	Y			Y			Y		Y	Y			Y	Y		Y		Y				Y					Y	
DIY	Y	Y	Y		Y	Y						Y				Y					Y	Y			Y		Y						Y		Y	Y
Computer games	Y		Y	Y			Y					Y	Y	Y		Y	Y		Y						Y		Y		Y		Y	Y		Y		Y
Sport fishing						Y			Y	Y						Y	Y								Y			Y								
Needlework												Y				Y					Y						Y								Y	
Theatre					Y				Y	Y		Y			Y								Y	Y			Y					Y			Y	
Pony trekking																Y			Y																	
Youth culture	Y			Y	Y	Y				Y		Y	Y			Y	Y	Y	Y						Y		Y	Y	Y	Y			Y	Y	Y	
Downhill skiing		Y				Y	Y		Y	Y			Y	Y		Y												Y								
Foreign policy		Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y		Y	Y	Y	Y			Y		Y	Y	Y		Y	Y		Y	Y	Y	Y	Y
Vocational training	Y		Y	Y	Y			Y	Y		Y	Y		Y	Y	Y		Y	Y		Y			Y			Y			Y	Y		Y	Y	Y	
Wine tasting		Y			Y			Y	Y	Y		Y	Y					Y	Y								Y	Y					Y			Y
Motorboats		Y	Y			Y			Y			Y				Y						Y		Y	Y		Y	Y	Y				Y			Y
Extreme sports	Y	Y			Y	Y			Y				Y			Y		Y	Y			Y	Y	Y			Y		Y	Y			Y	Y		
Closest MOSAIC segment based on combination of above tastes:	G23	B06	G23	G23	B08	G23	A01	A01	B06	B08	G23	B08	B08	A01	B08	G23	G23	B06	G23	G23	B06	G23	B08	B08	G23	B06	B08	G23	G23	G23	G23	D15	G23	G23	B08	B06

Experian Mosaic

A B C D E F G H I J K L M N

1 2 3 4

INSIGHTONE

Group A: **Affluent Pioneers**

Axel & Charlotte

Type A01: **Scandinavia's Wealthiest**

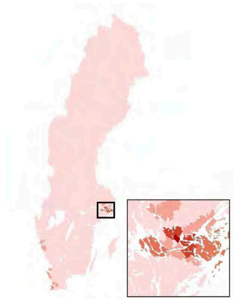
Families with extremely high incomes, in large houses in sought after suburbs close to larger cities

0.52% **0.70%**

Overview

Key Features

- Extremely wealthy
- Middle-aged couples
- Often have children living at home
- Houses outside large cities
- Have University degrees
- Trendy and fashion conscious
- Interested in art and culture
- High capital
- Extremely high purchasing power
- Adaptive to change



Rankings

Presence of children	4/44
Pensioners (65+)	39/44
Young (<26)	16/44
Household income	1/44
Interest expenses	2/44
Education level	2/44
Detached houses	12/44
Coop apartments	28/44
Private car	18/44
Company car	1/44

Typical Houses

Top Postal Areas

- 182 Djursholm/Stocksund
- 182 Danderyd
- 181 Lidingö Centrum Norr
- 167 Bromma, Alvik
- 429 Kullavik

If you're after a target group with strong purchasing power, you need look no further. This group is also interested in new and exciting products and services. Since they have just about everything in the way of material possessions, they're now after experiences. A golfing holiday to northern Portugal combined with tasting the local wines would be really great.

In the summer they enjoy coastal excursions along the Bohus coast in their First 45, and when winter comes they take their children to the ski slopes in Åre. There they meet up with their many neighbours and acquaintances.

On Sundays it's great to sink back in their Svenskt Tenn armchair and read Golf Digest, Allt om Vin or indeed the most recent issue of DI Weekend. They enjoy watching golf and cookery programmes and a good quality English detective series on their super-slim 50 inch TV.

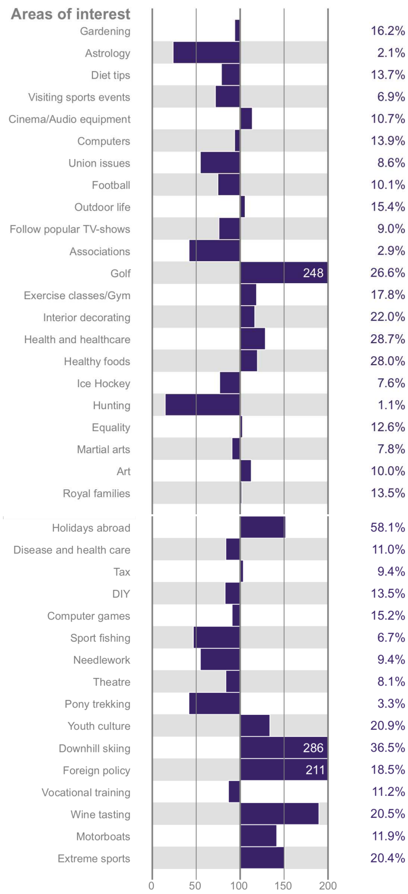
This group comes top in terms of education level, income and purchasing power. The children are starting to enter their teens and develop their own lives. This frees up more time for experiences, socialising with friends and travelling. They like pretty and expensive things, both for themselves and their nearest and dearest.

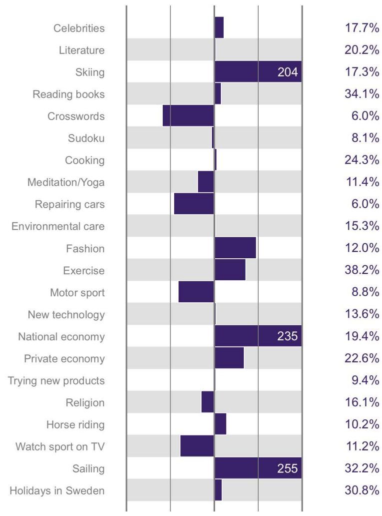
In Stockholm this group can be found in places such as Lidingö, Saltsjöbaden and Bromma. In Gothenburg they can be found in Askim or Kungsbacka. They often hold managerial or specialist posts in IT companies, consultancies, banking and finance or public administration. Their work requires them to stay abreast of things. To keep up socially as well; following what goes on at home, abroad and in other trends is essential.

Understanding Charts
Unless otherwise stated, charts show the Index and Mean %.

The Index is shown as a bar, and the Mean % is shown to the right:

Areas of interest





Experian Mosaic

A B C D E F G H I J K L M N

5 6 7 8 9

INSIGHTONE

Group B: **Metropolitan Pioneers**

Jacob & Ebba

Type B06: **Ethical City Families**

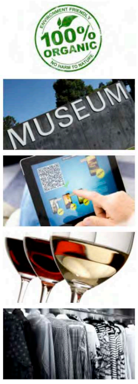
Urban families with high incomes, living in housing-association apartments, in the centre of larger cities

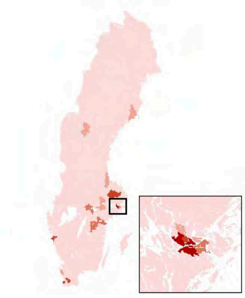
0.50% **0.59%**

Overview

Key Features

- Wealthy
- Middle-aged couples
- Often have children living at home
- Housing association flats in large city centres
- Have University degrees
- Trendy and fashion conscious
- Environmentally and socially conscious
- High purchasing power
- Fairly adaptive to change





Rankings

Presence of children	9/44
Pensioners (65+)	27/44
Young (<26)	38/44
Household income	5/44
Interest expenses	8/44
Education level	3/44
Detached houses	32/44
Coop apartments	9/44
Private car	39/44
Company car	6/44

Typical Houses

Top Postal Areas

- 114 Stockholm, Östermalm
- 111 Stockholm, City
- 111 Stockholm, Gamla stan
- 113 Stockholm, Birkastan
- 113 Stockholm, Vasastan/Döbelnsgatan

This is where you'll find individuals who are members of Fotografiska, read the SvD business pages and are interested in society at large, both in Sweden and abroad. They like experiences in the form of travelling to new and relatively unexplored places, preferably not going there by charter. Visiting a good restaurant or a trendy bar is essential. When they go shopping they like items that are a bit more trendy and expensive; smartphones and iPads, for instance, or perhaps an Omega or a Rolex.

The interests of this group reflect media consumption. This is where you'll find magazines such as Fokus, Filter and Elle on the coffee table. When they choose to spend time on the sofa in front of the TV, they like Kobra, Access or a documentary on Channel 4. Their taste in films is narrow, with a preference for independent productions.

This group is socially aware, trend-conscious and has strong purchasing power. On a Saturday morning, you'll find them at farmers' markets buying organic, trendy and locally baked sourdough bread. Being quality-aware is what counts.

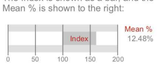
They live almost exclusively in the inner city, and an overwhelming number live in Stockholm, preferably in a turn-of-the-century flat with high ceilings. Many of them have children but have chosen not to move to a house as yet. A large proportion will never do so either; instead they love what the city has to offer together with its vibrant feel.

This is a group of high earners. An academic education goes without saying, and their work is a major and important part of their lives.

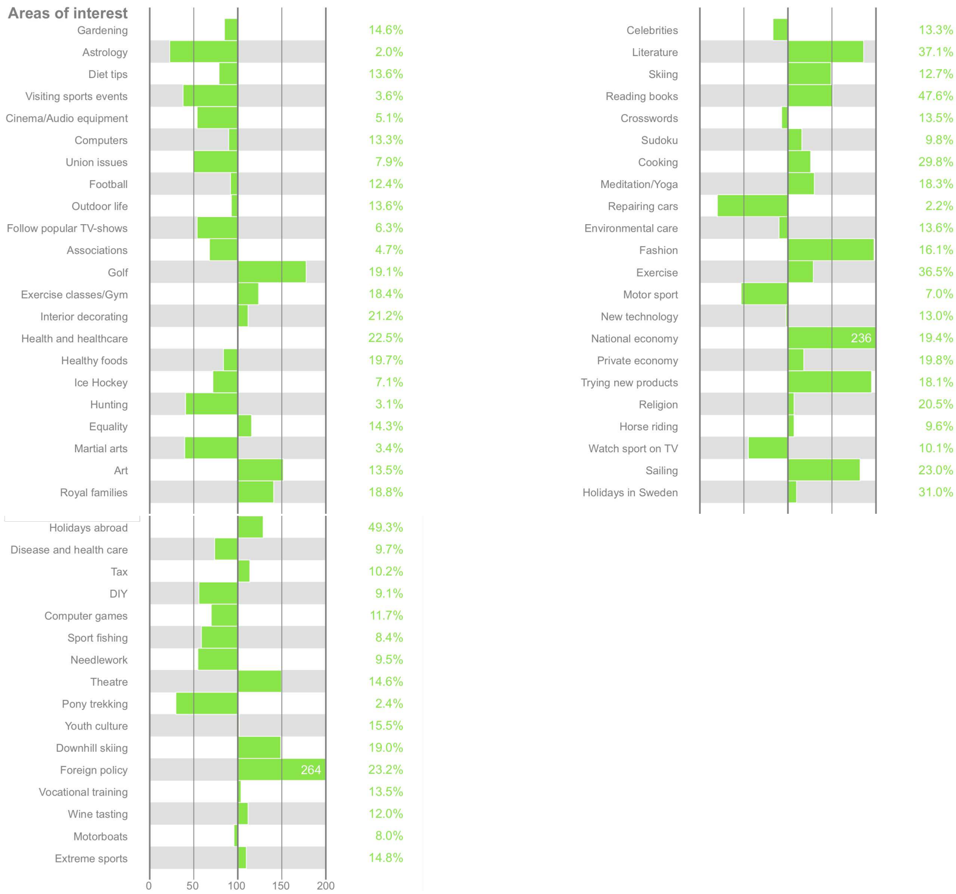
Typical occupations include management consultant, banking and finance specialist or media and communications project manager. Surprisingly, the proportion of this group that are second generation immigrants is sizeable.

Understanding Charts
Unless otherwise stated, charts show the Index and Mean %.

The Index is shown as a bar, and the Mean % is shown to the right:



Areas of interest



A B C D E F G H I J K L M N

5 6 7 8 9

Group B: **Metropolitan Pioneers**

Oscar & Sophie

Type B08: **Culture and Politics**

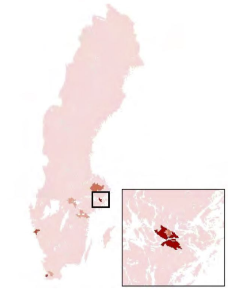
Urban singles with high incomes, living in apartments in sought after houses, in the centre of larger cities

2.05% **1.84%**

Overview

Key Features

- Singles
- High income
- Apartments in the centres of larger cities
- Have University degrees
- Interested in art and culture
- Trendy
- Environmentally and socially conscious
- Frequent users of social media
- Very high purchasing power
- Adaptive to change



Rankings

Presence of children	26/44
Pensioners (65+)	22/44
Young (<26)	25/44
Household income	23/44
Interest expenses	19/44
Education level	8/44
Detached houses	42/44
Coop apartments	10/44
Private car	41/44
Company car	10/44

Typical Houses

Top Postal Areas

- 111 Stockholm, Gamla stan
- 118 Stockholm, Södermalm mellersta
- 111 Stockholm, City
- 113 Stockholm, Norrmalm
- 413 Göteborg, Haga

This is where you'll find individuals who are members of Fotografiska, read the SvD business pages and are interested in society at large, both in Sweden and abroad. They like experiences in the form of travelling to new and relatively unexplored places, preferably not going there by charter. Visiting a good restaurant or a trendy bar is essential. When they go shopping they like items that are a bit more trendy and expensive; smartphones and iPads, for instance, or perhaps an Omega or a Rolex.

The interests of this group reflect media consumption. This is where you'll find magazines such as Fokus, Filter and Elle on the coffee table. When they choose to spend time on the sofa in front of the TV, they like Kobra, Access or a documentary on Channel 4. Their taste in films is narrow, with a preference for independent productions.

This group is socially aware, trend-conscious and has strong purchasing power. On a Saturday morning, you'll find them at farmers' markets buying organic, trendy and locally baked sourdough bread. Being quality-aware is what counts.

They live almost exclusively in the inner city, and an overwhelming number live in Stockholm, preferably in a turn-of-the-century flat with high ceilings. Many of them have children but have chosen not to move to a house as yet. A large proportion will never do so either; instead they love what the city has to offer together with its vibrant feel.

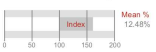
This is a group of high earners. An academic education goes without saying, and their work is a major and important part of their lives.

Typical occupations include management consultant, banking and finance specialist or media and communications project manager. Surprisingly, the proportion of this group that are second generation immigrants is sizeable.

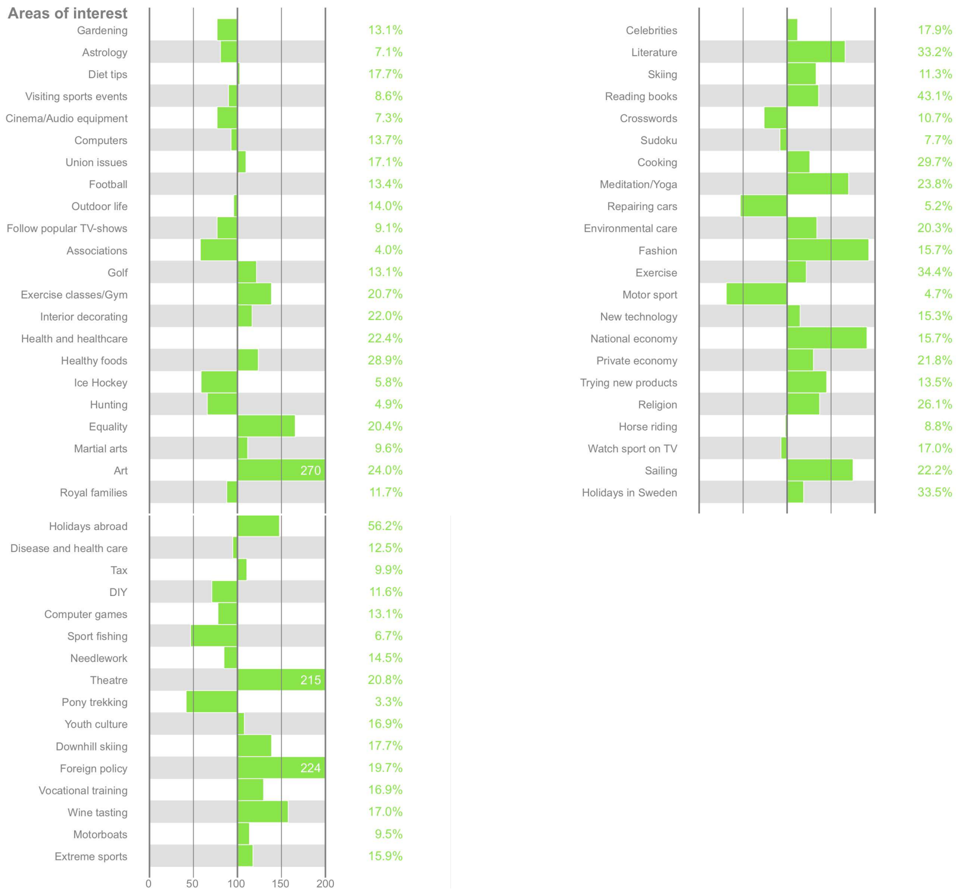
Understanding Charts

Unless otherwise stated, charts show the Index and Mean %.

The Index is shown as a bar, and the Mean % is shown to the right:



Areas of interest



Group D: **Curious Pioneers**

Emil & Lovisa

Type D15: **University Fringe**

Established students with low incomes and highly educated, in University towns

0.72% **0.62%**

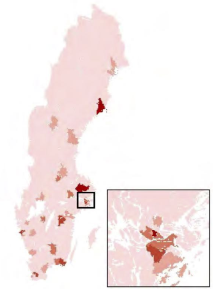
Overview

Key Features

- Higher education students
- Low income
- Low purchasing power
- Extremely adaptive to change
- High interest in new technology
- Likely to shop online
- Environmentally and socially conscious
- Frequent users of social media





Rankings

Presence of children	38/44
Pensioners (65+)	42/44
Young (<26)	2/44
Household income	43/44
Interest expenses	43/44
Education level	1/44
Detached houses	30/44
Coop apartments	15/44
Private car	43/44
Company car	41/44

Typical Houses


Top Postal Areas

- 115 Stockholm, Universitetet
- 907 Umeå, Ålidhem/Sofiehem
- 752 Uppsala Studentstaden
- 907 Umeå, Gimonäs/Carlshem
- 754 Uppsala Kapellgärdet / Kvarngärdet

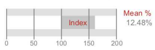
The people in this group are young and lead a hectic life with lots of interests. They are interested and engaged with the world and are connected in every sense. They maintain contact with their friends around the world via Facebook and Instagram. They might not have as much consumer power yet, but they buy a fair amount of clothes, video games and computers - maybe even a MacBook Air - often on hire purchase.

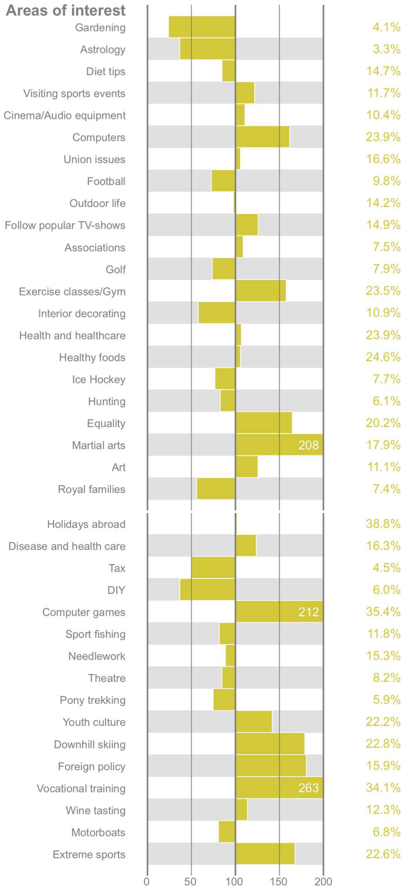
Naturally, they frequently go out to restaurants and bars; after all, they are students and single. The odd lazy Sunday sees them with their laptop on their lap enjoying a film from Netflix or music on Spotify, or watching the news on SVT Play. This group already watches TV in a way we will soon all be doing.

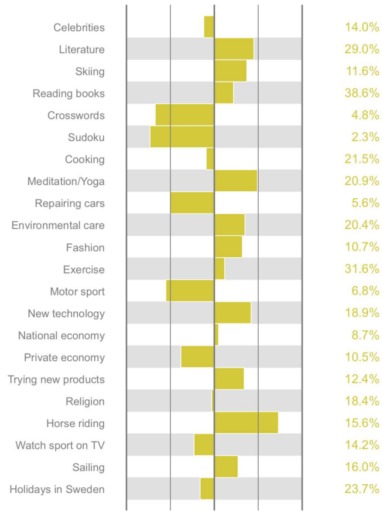
You'll find them in city suburbs such as Sundbyberg or Kortedala and university towns such as Lund, Uppsala and Linköping. They live in small rented flats and have not had children yet. Many of them take on extra work to eke out their study allowance. Those who are not students often work in hotels and restaurants or shops.

Understanding Charts
Unless otherwise stated, charts show the Index and Mean %.

The Index is shown as a bar, and the Mean % is shown to the right:







Experian Mosaic

A B C D E F G H I J K L M N

22 23

INSIGHTONE

Group G: **Multicultural Followers**

Omar & Fatima

Type G23: **Multicultural Families**

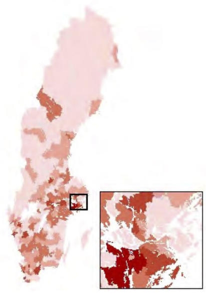
Young to middle-aged families, with low incomes, in rented flats in multicultural suburbs of large cities

2.83% 3.29%

Overview

Key Features

- Younger to middle-aged couples
- Often have children living at home
- Rental flats
- Live in multicultural suburbs of large cities
- Fairly adaptive to change
- Traditional
- Commonly have low credit ratings



Rankings

Presence of children	6/44
Pensioners (65+)	33/44
Young (<26)	5/44
Household income	40/44
Interest expenses	35/44
Education level	42/44
Detached houses	28/44
Coop apartments	19/44
Private car	36/44
Company car	42/44

Typical Houses

Top Postal Areas

- 163 Spånga, Rinkeby
- 151 Södertälje, Hovsjö
- 145 Fittja
- 213 Malmö, Rosengård
- 145 Norsborg

Many of the members of this group have foreign blood and grew up with influences from around the world. The multicultural element is a large one and is reflected in their interests and behaviour. Cross-country skiing is not the sport that instantly springs to mind as one to take up. Brazilian Capoeira or football would be more like it.

This group is also interested in religion and astrology. However, they're just as happy to relax with an hour or so of Assassin's Creed or FIFA 12 on their PS3 Console.

This group likes to buy DVD films, perfume or dieting products, preferably on their mobile. If you want to target them with advertising, you'll do best buying space in magazines such as PC för alla, Vi föräldrar or Foto. They enjoy family get-togethers over tapas, chicken khorma or meze, and watching soaps and TV series from their home country.

This group is young and lives in rental flats, reflecting the fact that they don't earn much money. Also, their education level is usually below the average. Many have not studied further than elementary school level.

The occupations they can be found in are service sector jobs in the hotel and restaurant industry, or else in health and social care. Many live in city suburbs such as Rinkeby, Rosengård or Bergsjön.

Understanding Charts
Unless otherwise stated, charts show the Index and Mean %.

The Index is shown as a bar, and the Mean % is shown to the right:

Areas of interest